ARQ 21 WINTER 2000

THE JOURNAL OF THE DEFENSE ACQUISITION UNIVERSITY

ACQUISITION Review QUARTERLY

VOL. 7 NO. 1

The ***Acquisition Review Quarterly*** is published quarterly for the Defense Acquisition University by the Defense Systems Management College Press, 9820 Belvoir Road, Suite 3, Fort Belvoir, VA 22060-5565. Periodicals Postage Paid at Fort Belvoir, VA and at additional mailing offices. Postmaster send changes of address to: Editor, ***Acquisition Review Quarterly***, Defense Systems Management College Press, 9820 Belvoir Road, Suite 3, Fort Belvoir, VA 22060-5565. For free copies, submit written requests to the above address. Articles represent the views of the authors and do not necessarily reflect the opinion of the Defense Acquisition University or the Department of Defense. ISSN 1087-3112.

The ***ARQ*** is available electronically on the DSMC Home Page at **http://www.dsmc.dsm.mil**

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE Winter 2000	3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE

Acquisition Review Quarterly (ARQ)
Vol. 7, No. 1

5. FUNDING NUMBERS

6. AUTHOR(S)

Numerous Authors

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Defense Systems Management College
Attn DSMC Press
9820 Belvoir Road Ste 3
Ft. Belvoir VA 22060-5565

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Defense Acquisition University
2001 N Beauregard Street
Alexandria, VA 22311

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The primary goal of the Acquisition Review Quarterly (ARQ) is to provide practicing acquisition professionals with relevant management tools and information based on recent advances in policy, management theory, and research. ARQ addresses the needs of professionals across the full spectrum of defense acquisition, and is intended to serve as a mechanism for fostering and disseminating scholarly research on acquisition issues, for exchanging opinions, for communicating policy decisions, and for maintaining a high level awareness regarding acquisition management philosophies.

14. SUBJECT TERMS
Engineering Mgmt.; Performance Measures; Distributed-Simulation; Acquisition Reform; Systemic Fiscal Optimism; Management

15. NUMBER OF PAGES
100

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

ACQUISITION Review QUARTERLY

Winter 2000
Vol. 7, No. 1

TABLE OF CONTENTS

ENGINEERING MANAGEMENT TRAINING: *COMPARING EXPERIENTIAL VERSUS LECTURE METHODS OF INSTRUCTION*

Robert H. Lightsey, D. Sc.

While many studies have compared passive and active instructional methods, none provides statistical evidence that one method is clearly superior. When the subject matter to be taught is technical in nature, however, the experiential method has been shown to be more effective in terms of both student reactions and learning.

The environment faced by most organizations today is characterized by increasing dependence on technology, whether that technology is a product of the organization or the tool it uses to compete in the marketplace. In addition to the emergence of technology as a dominant consideration in the strategy of organizations, the pace of technological change has accelerated to the point that technical obsolescence is a concern that affects both products and people. In 1994 the Motorola Corporation estimated that the knowledge of the average engineer becomes obsolete in 2 to 5 years (Motorola, 1995).

Furthermore, competition has become global in nature, increasing the pressures on organizations to be both technically agile and to be economically efficient producers. In this environment of complex technologies, rapid technical obsolescence, and global competition, organizations have increasingly accepted the idea that education and training are key to the ability of the

organization to compete in the global economic environment.

Training is a major commitment for technically focused firms today, and it has become a big business. Motorola, for example, affords each of its 132,000 employees one week of training each year (Motorola, 1995), and, according to one estimate (Dipoye, Smith, and Howell, 1994), U.S. corporations having 100 or more employees spent $43.2 billion for training during 1991. With this level of commitment in mind, it becomes particularly important that training be conducted as effectively as possible, and that the instructional methods used be those that contribute most to improving the job performance of the student.

This article describes a study that made a side-by-side comparison of instructional methods used to teach two large groups of students attending the Advanced Program Management Course (APMC) at the Defense Systems Management College (DSMC). One group was taught the systems engineering management portion of the course using lecture and discussion as the primary teaching method, while the second group was taught experientially using a hands-on design project. The learning objectives of the two courses were identical and the student groups evaluated were very homogeneous. This appeared to be an excellent opportunity to conduct a controlled study in order to investigate the impact of teaching methods on learning outcomes. Given the substantial commitment made by business and government to education and training, the results of this investigation could be important in determining the nature of training for students in future courses.

"Training is a major commitment for technically focused firms today, and it has become a big business."

RELATED LITERATURE

Perhaps the best known and most comprehensive approach to the assessment of training was developed by Kirkpatrick (1977). His model comprised four discrete and progressive levels of evaluation: student reactions, learning achieved, transfer of changed behaviors to the workplace, and results achieved in the workplace. Most evaluation does not go beyond assessment of student reactions (Dipoye, Smith, and Howell, 1994) and appears to be based on the assumption that, if the student leaves the course with a positive attitude regarding the training, then there will be positive results in other measures of effectiveness.

Research indicates that this assumption is likely to be a poor one. In a study evaluating the relationships among the different levels in the Kirkpatrick model, Alliger and Janak (1989) found no significant relationship between student reactions to training and the higher levels—learning, transfer to the workplace, or results achieved. This makes intuitive sense; the course rated favorably by students may not be the one that provides useful learning. On the other hand, Alliger and Janak found a positive relationship among the higher levels, indicating that learning, once achieved, will likely result in the transfer of lessons learned to the workplace, and subsequently to improved results in the workplace.

The fact that there is positive linkage among the higher three levels, but no relationship between student reaction and learning, suggests that learning is the critical indicator of course effectiveness; learning is the linchpin. If the student learns, the course is likely to result in a positive change at all levels of effectiveness. Others, such as Landy (1987) and Maier (1973), have observed that, while the prevailing assumption has been that attitudes influence behavior, for this to be true requires a confluence of other factors, such as experience and motivation. Evaluation that stops at the assessment of student reaction too often provides little in the way of useful information regarding the probability that the training results in learning or that the training will carry over into the workplace.

Adult learning theory, as espoused by Knowles (1980) and others, implies that adults will learn and retain more when they take an active role, participate more, and use multiple senses in the learning process. A number of research efforts have been conducted with the objective of evaluating the extent to which teaching methods influence learning, but few statistically significant conclusions have been recorded. According to Rachal (1994), "...advocacy of andragogy as a superior strategy for facilitating adult learning does not seem to be borne out by the existing empirical studies... ."

Table 1 summarizes a number of studies that have compared teaching methods. All compared various participative and experiential methods with a control group taught using more passive methods. Significantly, none established a statistical difference in the learning achieved using one method in preference to the other. Campbell noted in his later work (1988) that analyses performed in this topic area are increasingly well structured and rigorous; however, even in more recent studies, there is little statistical evidence that experiential methods are superior to more passive methods.

"If the student learns, the course is likely to result in a positive change at all levels of effectiveness."

In spite of this consistent pattern of failure to find differences in the results achieved between instructional methods in various types of training courses, the present study was structured specifically to compare the effectiveness of alternative methods. Many of the studies reviewed seemed to have one or more problems:

- In some cases the differentiation between "participative" and "nonparticipative" conditions was nebulous.

- In many cases the samples were quite small, so the power to discriminate was lessened.

- Where the samples were reasonably large, the research was conducted in environments (universities, corporations) that are apt to be subject to a great deal of outside interference (Walleri and Japely, 1986).

The numbers of subjects available at DSMC, the homogeneity of the groups who attend courses (Table 2), and the fact that students are largely isolated from their work and outside interference suggested that the environment would be ideal for a

Table 1. Related Studies of Comparative Teaching Methods

Author	Group	N	Methods Compared	Instruments	Analytic Methods	Results (Learning)
Bretz and Thompsett (1992)	Kodak Corp. MRP Training	180	Lecture vs. integration	Survey, pretest, learning	ANOVA Correlation posttest	No difference $P = .01$
Carr (1982)	University (Economics)	26	Lecture vs. case study	Pretest, posttest	T-test, ANOVA	No difference $P = .05$
Carter (1995)	University (physical training)	36	Lecture vs. case study	Survey, posttest (only)	MANOVA	No difference $P = .10$
James (1991)	Adult (education)	31	Lecture vs. case study	Pretest, posttest	ANOVA T-test Regression	No difference $P = .05$
Merrill (1995)	Adult medical (cardiac)	37	Lecture vs. self-study	Pretest, posttest	ANCOVA	No difference $P = .05$
Thoms and Klein (1994)	Adult (hospital management)	64	Nonparticipation vs. participation	Survey, multiple tests	ANOVA Correlation Chi-Square	No difference $r = .16$
Ward (1993)	Navy (medical)	300	Nonparticipation vs. participation	Survey, posttest (only)	T-test, ANOVA	Learning not assessed
Welch (1990)	University students (business)	181	Lecture vs. active methods	Pretest, posttest	ANOVA	No difference $P = .05$
White (1995)	University students (tech)	112	Lecture vs. computer-aided training	Survey, pretest, posttest	ANOVA	No difference $P = .10$

comparison of teaching methods in conditions where those methods were likely to be the primary determinant of differences observed in performance. This is an issue of particular interest since many organizations are moving rapidly to implement curricula that are experiential and hands-on (Parkinson, 1994; Raelin and LeBien, 1993) rather than lecture-based and theoretical, in spite of the lack of empirical research findings that support that trend.

METHOD

SUBJECTS AND METHODS COMPARED

The APMC is a training course for Department of Defense (DoD) program managers, conducted at DSMC. Those attending the course are selected from that portion of the DoD acquisition workforce who either hold senior management positions (program managers, deputy program managers, functional managers, and division heads) in program offices, or who are being prepared for such positions.

While they take the course, they are relieved from their assigned jobs, are relocated to DSMC, and are expected to attend classes daily in various topics related to project management.

APMC 95-1 was taught the engineering management portion of the course using lecture-discussion methods with limited exercises. APMC 96-1 entered a year later, but, rather than lecture, this group was taught by integrating the instruction with a design project. The students in APMC 96-1 were required to plan, design, build, and then test a vehicle based on a set of performance requirements issued at the beginning of the course. All engineering management instruction was woven into, and was related to, the design project. APMC 95-1 was the control group, and APMC 96-1 was the treatment group for purposes of this study.

Each offering of the APMC included approximately 420 students—a representative mix of military and civilian employees in DoD research and development

Table 2. Demographics of Control and Treatment Groups

Attribute	APMC 95-1	APMC 96-1
Affiliation		
Air Force	29.7%	30.5%
Army	28.9%	25.0%
Navy/Marines	28.6%	32.9%
Other	12.9%	11.6%
Highest education level		
Masters/Ph.D.	69.4%	71.4%
Bachelors/other	30.6%	28.6%
Education type		
Technical	48.6%	47.1%
Management/other	51.4%	52.9%
Military	55.3%	50.7%
Civilian	44.7%	49.3%
Female attendees	13.6%	16.9%
Acquisition experience (years)	9.5	10.8
DoD experience (years)	15.5	17.4
Sample total	360[a]	420

[a] Two sections (60 students) were excluded from the APMC 95-1 control group, because they were used to pilot the techniques that were later used in teaching the APMC 96-1 course.

activities. In addition, a few employees of defense industry corporations typically attend each class offering. Table 2 compares the demographics of the two groups involved in this study. The homogeneity of the students who attend these courses is remarkable—they are of similar ages (typically 35 to 40 years old); they tend to have similar educational and experience backgrounds, and they are similar in civilian/military mix as well as in service affiliation.

Each class was divided into 14 sections of 30 students. Section assignment was a stratified random process; subgroups were established proportionally by service affiliation (Army, Navy, Air Force), then assignment to sections was random. The result was a distribution that was proportionally representative of the entire class in terms of service or industry affiliation, and which was random in terms of assignment of individuals to sections.

Sections received the same program of instruction based on an established curriculum with learning objectives that were identical both among sections and between the two class groups. Each section was taught by an assigned instructor. The 12 sections in the control group were taught by eight instructors (four of whom taught two sections); the 14 sections of the treatment group were taught by 12 instructors (5 of whom had taught the control group). Neither instructors nor students were aware that student data would be analyzed for the purposes of this study. A concern that the control group (95-1) might have treated the course with less seriousness (since this was the last offering of the format in use at the time) were resolved by comparing the end of course comprehensive exam scores of APMC 95-1 with previous classes. Their scores were essentially equal to those of their predecessors, indicating no lessening of effort or learning on their part.

INSTRUMENTS

This study evaluated the effectiveness of the training offered in terms of the first

1. Are you employed as a
   a. Government civilian
   b. Military
   c. Contractor
   d. None of the above

2. Highest level of formal education achieved
   a. Bachelor's degree
   b. Master's degree
   c. Doctoral/Ph.D.
   d. Other (specify)________________

Figure 1. Sample Items from Pretest Demographic Survey

two levels of the Kirkpatrick model, and consequently measurement of student reactions and learning were key. The collection of data was handled through administration of pretest and posttest instruments.

Pretest. The pretest included a test of knowledge in the engineering management domain and a survey of demographic factors. The domain knowledge test is discussed in more detail below. The demographic questionnaire surveyed employment and educational background, experience, and previous training prior to exposure to the APMC course. An example of two items on the demographic survey is shown in Figure 1.

Posttest. The posttest consisted of a questionnaire that measured student reactions to the course and a second test of knowledge in engineering management. Both were administered at the final training session. The questionnaire included five questions and also provided for subjective student comments. Three of the questions addressed the extent to which students found the course of instruction informative, enjoyable, and characterized by reasonable workloads; these were the questions viewed to best reflect student reactions to the course, as described by Kirkpatrick. The remaining items dealt with preferences for hands-on versus lecture and the future usefulness of the training in the work environment. Responses were given in numerical form using a seven-point Likert scale. Figure 2 shows a sample question.

Domain knowledge test. The domain knowledge test was the means for measuring learning. A panel of subject matter experts in engineering management from DSMC evaluated all questions proposed for content validity. This panel consisted of six experienced systems engineering managers who were teaching in the APMC, all of whom were Level III members of the Defense Acquisition Work Force. Questions were formally scored on the extent to which they satisfactorily addressed the learning objectives with which they were associated. These learning objectives were derived from the set of competencies required of the systems engineering management subcourse in the APMC.

The internal reliability of the instruments was evaluated over a period of several months by testing them on several groups of students who were attending

I found the systems engineering course material informative.

Strongly disagree   Strongly agree

|——|——|——|——|——|——|

1   4   7

Figure 2. Sample Item from Posttest Student Reaction Questionnaire

DSMC short courses in systems engineering management. In terms of experience levels and other key variables such as educational backgrounds, the attendees at Level III short courses are comparable to the attendees at APMC. The major difference is that the attendees at the specialized short courses in engineering management tend to be predominantly from the technical disciplines and to hold technical management positions, while the APMC includes a more balanced mix of managers from various functional areas.

Tests were administered using the split-half procedure. As experience was gained through repeated testing in succeeding courses, adjustments were made in the content and mix of the tests based on item analyses. The reliability was finally demonstrated to be 0.930. As an additional check, the entire question set was also administered as a single exam, and the reliability was calculated using the Kuder-Richardson (KR 20) formula. The results of that calculation indicated that the mean reliability of all split half combinations is 0.83, still well within the bounds normal to skills and competency based tests in use.

The full question set was divided into halves with demonstrated reliability 0.93. In their final form the two tests (pretest and posttest) were each 15 questions that tested student knowledge at the application level and which had been demonstrated to do so equally. An example from one of those tests is shown in Figure 3.

Procedure. The pretest instrument was administered to both the control and treatment groups during their first lesson in the systems engineering subcourse. The posttest was administered to both groups

As your program approaches CDR, your systems engineer informs you that, in his/her estimation, the design is about 60% complete. Assuming that his/her estimate is correct, you should (choose one):

a. Continue as planned and hold the CDR as scheduled. In today's environment, design maturity is not an issue for government program managers.

b. Hold the CDR as scheduled. There are minimal risks associated with early CDRs, since designs continue to mature until well after CDR is completed.

c. Delay holding the CDR until the design is substantially complete. A design should be 85–90% complete before the CDR is conducted.

d. Delay the CDR until the contractor completes and delivers the system specification for government review.

Figure 3. Sample Domain Knowledge Test Item

Table 3. Two-Sample *t* Test for Pretest[a]

Class	N	Mean	SD	t
APMC 95-1	359	57.2	15.9	1.02[b]
APMC 96-1	407	58.4	16.2	

[a] H_0: APMC 95-1 = APMC 96-1.
[b] $P = .30$.

at the end of that course. Taken together, the two provided the means to compare knowledge levels between and within groups, both before and after exposure to the engineering management course of training. They also provided the means to relate knowledge of engineering management concepts and principles to demographic factors, and, finally, they measured student reactions to the training received. The administration of these two instruments completed the data collection for the conclusions drawn here.

This was a pretest-posttest control group experimental study. The span of time between the pretest and posttest administration was 10 weeks. Pretest-posttest interaction was controlled in this case by the timing of the two tests and by using different questions in the pretest and posttest. The differences in the pretest and posttest, in combination with a span of more than two months between measurements, was considered adequate to guard against pretest-posttest interaction.

Results of the Analyses

Initial Knowledge Levels

While the data in Table 2 indicate strong similarity between the groups, it was important to determine whether or not the two sample groups were equal with respect to the level of knowledge upon entry to the course. The results (Table 3) indicate that the initial knowledge level, as shown by mean pretest scores of the two groups, was equal (P =?.05). Initial equality between groups made more supportable a conclusion that the differences observed later could be attributed to exposure to the training, rather than to differences that may have existed at the outset of the experiment.

The analysis then turned to investigation of hypotheses that were structured to parallel the first two levels of the Kirkpatrick model (1977). These were that students would react more favorably to a curriculum taught using experiential methods, and that students would learn more from curriculum taught using experiential methods.

Student Reactions

Student reaction data were collected using the posttest questionnaire described earlier. The three variables used as indicators of student reactions and associated statements were:

- Informed: I found the engineering management course informative.

- Enjoyed: I enjoyed the engineering management course.

- Workload: The course covered too much material for the time allotted.

Analytic results indicate that APMC 96-1, the class taught using experiential methods, believed that the class was more informative, enjoyed the course more, and perceived the workload to be less oppressive (Table 4). This last finding is interesting in that the requirement to design, fabricate, and test vehicles appeared to have represented a considerably increased workload over that associated with the course taught using lecture-discussion methods, yet the students indicated that they perceived the workload to be less.

Narrative reactions were also solicited from the students of the two courses. The results, while more subjective, corroborated the analytic results of the t tests in that the comments of APMC 95-1 indicated less satisfaction than did APMC 96-1.

LEARNING

A number of studies (e.g., Bretz and Thompsett, 1992) have concluded that exposure to experiential instruction will result in student reactions that are more positive than those elicited from classes exposed to the more traditional lecture and discussion approach to teaching. The results reported above confirm and strengthen these findings.

Theory leads one to expect that the experiential methods would produce superior results, not only in student reactions, but also in learning. To measure learning, this study used the combination of the

Table 4. Two Sample t Test for Student Reactions

Variable	N	Mean (Scale of 1 to 7)	SD	t
Informed				
APMC 95-1	321	4.67	1.30	–4.12[a]
APMC 96-1	393	5.12	1.57	
Enjoyed				
APMC 95-1	322	4.51	1.41	–5.01[a]
APMC 96-1	393	5.09	1.64	
Workload				
APMC 95-1	322	3.41	1.48	4.78[a]
APMC 96-1	392	2.88	1.44	

[a] $P = .01$.

Table 5. Student Comments on Engineering Management Course

Category	APMC 95-1 (%)	APMC 96-1 (%)
Satisfied (commented positively)	10	14
Dissatisfied (commented negatively)	17	15
Satisfied/neutral (commented generally, but neither favorably nor unfavorably)	13	17
Neutral (did not comment)	60	54

pretest and posttest to measure knowledge levels at the beginning of training and at the completion. Analysis of covariance was used to adjust observed posttest scores, controlling for differences observed in pretest results. We used *t* tests to measure both the extent to which each instructional method resulted in positive learning (pretest vs. posttest) and to measure the extent to which one method produced more learning than the other (posttest vs. posttest). Table 6 gives the results of those tests (pretest scores are repeated for convenience).

These tests indicate that both lecture and experiential methods produced positive learning. The tests furthermore indicate that the experiential method produced a higher level of knowledge as indicated by posttest scores. This statistically supports the theoretical hypothesis that experiential training produces an improved level of learning; however, the differences measured were smaller than might have been expected or hoped for. More will be said on this topic later.

Relationship between Student Reactions and Learning

Alliger and Janak (1989) found no significant relationship between student reactions and learning. This study presented an opportunity to evaluate whether or not that same lack of relationship was in evidence in this study, as well. Since the demographics (Table 2) and initial analysis of pretest scores indicated that the two groups involved could reasonably be considered to represent a single underlying population, the two were combined into a single large group for purposes of this analysis.

Table 6. Domain Knowledge Test Scores

Group	Pretest Mean	Posttest Mean	Posttest Mean (Adj.)	Posttest SD	*t* Test (Pre-Post)	*t* Test (Post-Post)
APMC 95-1	57.2	73.2	73.4	16.6	–12.67[a]	–3.93[a]
APMC 96-1	58.4	77.8	77.7	14.1	–17.89[a]	
[a] $P = .01$.						

Table 7.
Multiple Regression—Student Reactions and Pretest on Posttest Score[a]

Predictors	Dependent Variable, Posttest Score				
	ß	SD	Seq. SS	*t*	*P*
Constant	64.332	3.367		19.1	0.000
Informed	–0.594	0.849	188.6	–0.70	0.484
Enjoyed	0.360	0.800	5,13.7	0.45	0.653
Workload	–1.754	0.392	6,596.3	–4.48	0.000
Pretest	0.301	0.037	16,374.1	8.34	0.001

[a] R-Sq (Adj.) = 11.9%.

Using the variables "enjoyed," "informed," and "workload" (described earlier) as the indicators of student reactions, we performed a multiple regression on posttest score. While "workload" was significant ($P = .05$), these variables together explained little of the total variance in posttest scores, leading to the conclusion that student reactions are essentially unrelated to learning, and thus supporting the Alliger and Janak observation.

As a further check, pretest scores were then added as an independent variable. Table 7 shows the result of the regression analysis. While the addition of pretest scores improved the model somewhat, the coefficient of determination remained quite low, indicating that the outcome of the course of instruction was primarily explained by factors other than either pretesting or student reactions. There was a moderate correlation ($r = .33$) between pretest and posttest scores, i.e., those who score well on the pretest tend to score well on the posttest and vice versa.

Other Factors that Influence Initial Levels of Knowledge

The pretest instrument included a survey that requested information regarding the student's job status, educational level, educational specialty, previous work experience, and professional training. The final part of the study involved analysis of the extent to which pretest score was influenced by these categorical factors. Pretest scores were used as an indicator of typical knowledge levels among the acquisition workforce population at large. The data were grouped to enable analysis of factors, each with multiple states as shown in Table 8. Factorial ANOVA was used for the analysis.

Education type, years of program management experience, and the nature of training courses taken previously were most significant. Somewhat surprising was the finding that the nature of current job held (technical or nontechnical) and the interval since attendance at training courses did not produce significant differences.

Table 8. Effect of Categorical Factors on Pretest Performance

Factor	States Addressed	Adj. SS	F	*P*
Employment status	Military or civilian	121.1	0.52	0.470
Education level	Graduate versus other	968.3	4.18	0.041
Type of education	Technical versus nontechnical	4,313.9	18.61	0.000
Current job	Technical versus nontechnical	114.6	0.49	0.482
Management experience	Years of experience	2,581.9	11.14	0.001
Gender	Male or female	672.1	2.90	0.089
Previous training courses	None, general, or functionally specific	3,338.5	14.41	0.000
Interval	Years since last training	14.9	0.06	0.800
Error	*df* = 693	160,602.5		

Since the type of educational background (technical or nontechnical) that the individual possessed produced very significant differences in test scores, it might have been expected that the type of job currently held would produce similar differences. Formal technical education appears to convey a broader, more integrative level of knowledge that contributes to the individual's ability to organize and control resources to accomplish a technical objective (engineering management), while holding a technical job did not. The nature of technical work is often highly specialized and relatively narrowly focused, which may explain the fact that those with technical jobs did not score significantly better on the pretest.

The nature of previous training courses attended deserves comment, also. The prerequisite training courses attended by students at APMC tend to fall into two general categories: those that address program management in general by exposing the student to relatively short subcourses in a variety of disciplines (of which engineering management is one), and those that are focused on a specific management discipline (such as engineering management or contract management).

This study found that students exposed to multidisciplinary general management courses did not perform significantly differently on the engineering management pretest from those who had not attended previous training. On the other hand, those who had attended training that was functionally specialized to teach engineering management performed significantly better on the pretest than all

other subgroups, including those who had attended general management courses and those who had attended no training at all.

In addition to these primary effects, the interactions among factors provided several interesting insights. For example, the results indicated that technical management experience can go far to overcome any disadvantages associated with nontechnical formal education. The relative advantages (in terms of pretest score) associated with technical education are largely isolated to those who have little or no experience. When both had experience in program management offices, technically educated and nontechnically educated students essentially performed equally.

"This study found the differences in learning produced by the two methods investigated to be smaller than expected...."

In this sense, education can be viewed as a means of achieving the equivalent of a level of experience prior to actually entering the technical management environment. Similarly, the study did not indicate that the time interval elapsed since training was influential in determining pretest scores. When the interaction of the elapsed interval and experience was evaluated, the results indicated that experience more than accounted for any losses due to the passage of time since training.

Discussion and Conclusions

These data support the hypothesis that students react more favorably to instruction that uses more experiential, active approaches to learning. While this and more traditional approaches to teaching result in positive learning experiences for the student, the work documented here further indicates that students learn more when experiential teaching methods are used in training. These findings in combination should encourage designers of technical training to structure courses that feature hands-on learning where theoretical and conceptual topics are reinforced by experiences gained during the training. Too many courses relegate applications and lessons learned through experience to the post-training period.

This study found the differences in learning produced by the two methods investigated to be smaller than expected; other studies have found no conclusive differences at all. There is a possibility that the problem has been in the instruments used to measure learning. When an identical instrument is used to measure knowledge levels between groups (as was the case in this study where the same pretests and posttests were used for each of the two groups), comparisons are naturally restricted to the knowledge levels measured by the instrument. In the case of most objective tests, knowledge measured is limited to levels two or three of the Bloom taxonomy (1956).

It is possible that the instruments used in this study did not capture the full extent of learning in the experiential course of training. For example, students synthesized designs, evaluated them, fabricated models based on the designs, and then evaluated the extent to which the products met original requirements. The tests used, however, did not address the higher levels of learning associated with this activity. This study was able to demonstrate

statistical differences between methods, due in some part to the large size of the control and treatment groups involved (nominally 360 and 420).

Future studies should take care to make certain that measurements of results fully account for all of the learning involved. If they do, more conclusive support for the use of experiential methods may be found. Additional research in this area would be useful, since the current literature is largely inconclusive on the topic of differences in learning outcomes as a function of instructional methods used.

This research also supports earlier work that found little or no relationship between student reactions to training and the learning that students achieve. Few organizations evaluate training at all; and many courses evaluate training only at the level of student reactions, assuming that positive reactions will result in positive learning and transfer of training to the workplace. That is apparently not a well-justified assumption. The evidence is conclusive that training evaluation must include an assessment of learning to be of value as an indicator of training effectiveness. Training is too costly and too important to the future of both organizations and their people to tolerate less.

Finally, this study addressed the extent to which various factors influence technical management knowledge levels in general. Education type (technical or otherwise), previous management experience, and the nature of previous training appear to be most influential. Surprisingly, neither the interval since most recent training nor the nature of the job currently held appeared to have substantial impact on knowledge levels. This has a number of interesting implications. For example, engineers generally need additional education or management experience to be good technical managers; they do not acquire technical management knowledge naturally as a by-product of their technical expertise. Another observation was that training, to be most effective, needs to be functionally focused. Courses that address too broad an array of functional topics will likely prove ineffective at training in any of them.

These findings must be balanced against certain limitations, including those associated with the instruments used to evaluate the learning achieved through the use of experiential methods. It is possible that the results documented did not account for all the learning achieved.

Another consideration is that the domain knowledge investigated is specific to the engineering management models that are common to the Department of Defense and to much of the U.S. aerospace industry as reflected in current industry standards, such as the Electronic Industries Association (EIA, 1994) and Institute of Electrical and Electronics Engineers (IEEE, 1994) standards for systems engineering management. But the findings may not be applicable to nontechnical fields or to environments in which the standard models for engineering management do not apply.

"Another observation was that training, to be most effective, needs to be functionally focused."

In spite of these limitations, this work extends previous analytic work and

confirms theoretical work in an area where empirical results are sparse. It also illuminates several areas—associated with training individuals to perform as engineering managers—that merit further study.

Dr. Robert H. Lightsey is chairman of the Systems Engineering Department at DSMC where he has taught for 7 years. He is a retired Air Force colonel whose career included substantial experience in both operations and acquisition. Before joining the DSMC, Lightsey worked for 5 years in the private sector. He holds a D.Sc. degree in engineering management from The George Washington University.

(E-mail address: lightsey_bob@dsmc.dsm.mil)

References

Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology, 42,* 331–342.

Bloom, B. S. (1956). Taxonomy of educational objectives: The classification of educational objectives (Book 1). *The cognitive domain.* New York: Longman.

Bretz, R. D. Jr., & Thompsett, R. E. (1992). Comparing traditional and integrative learning methods in organizational training programs. *Journal of Applied Psychology, 77*(6), 941–951.

Campbell, J. P. (1988). Training design for performance improvement. In J. P. Campbell and R. J. Campbell (Ed.), *Productivity in organizations.* San Francisco: Jossey-Bass Publishers.

Carr, N. J. (1982). An evaluation of the use of case studies in economics at advanced levels in schools. In *Research papers in economics education.* London: London University, Institute of Education.

Carter, L. D. (1995). *Effectiveness of case-based method versus traditional lecture in the retention of athletic training knowledge.* Doctoral dissertation, Oregon State University.

Dipoye, R. L., Smith, C. S., & Howell, W. C. (1994). *Understanding industrial and organizational psychology: An integrated approach.* Fort Worth, TX: Harcourt Brace College Publishers.

Electronic Industries Association. (1994). Interim Standard 632, Systems Engineering. Washington, DC: EIA Engineering Publications Office.

Institute of Electrical and Electronics Engineers. (1994). IEEE 1220: Standard for Application and Management of the Systems Engineering Process. New York: IEEE Standards Department.

James, F. R. (1991). An analysis of case-based instruction in teacher preparation classroom management training. Doctoral dissertation, University of Virginia.

Kirkpatrick, D. L. (1977). Evaluating training programs: Evidence versus proof. *Training and Development, 31,* 9–12.

Knowles, M. S. (1980). *The modern practice of adult education: From pedagogy to andragogy.* New York: Cambridge Co.

Landy, F. J. (1987). *Psychology: Science of people.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Maier, N. F. (1973). *Psychology in industrial organizations.* Boston: Houghton Mifflin.

Merrill, P. A. (1995). *A comparison of two teaching methods in basic cardiac life support training and education.* Master's thesis, Grand Valley State University.

Motorola Corporation. (1995). *1994 Annual Report.* Schaumburg, IL: Motorola Corporation Corporate Offices.

Parkinson, G. (1994, October). Hands on learning: The new wave in Ch.E. education. *Chemical Engineering,* 45–49.

Rachal, J. R. (1994). Andragogical and pedagogical methods compared: A review of the experimental literature. Paper published by U.S. Department of Education, Office of Educational Research and Improvement, Washington, DC.

Raelin, J. A., & LeBien, M. (1993). Learn by doing. *HR Magazine,* 61–70, February.

Thoms, P., & Klein, H. J. (1994). Participative and evaluative outcomes in management training. *Human Resources Development Quarterly,* 5, (Spring): 27–39.

Ward, F. P. (1993) *An investigation of the effects instructional strategies have on student attitudes.* Doctoral dissertation, The George Washington University.

Walleri, R. D., & Japely, S. M. (1986). *Student intent, persistence, and outcomes.* Paper presented at the 26th Annual Forum of the Association for Institutional Research, Orlando, FL.

Welch, B. D. (1990). *Effects of method of instruction on student achievement, student description of teaching, and the interaction of student motivation, cognition strategies, and resource management strategies in an undergraduate business class.* Doctoral dissertation, Texas A&M University.

White, M. J. (1995). *Relationship between teaching method and student achievement in technical college introduction to microcomputer courses.* Doctoral dissertation, Auburn University.

TECHNICAL PERFORMANCE MEASURES AND DISTRIBUTED-SIMULATION TRAINING SYSTEMS

Dr. Michael Proctor and MAJ Michael J. Lipinski, U.S. Army

Simulation systems are being increasingly used as a cheaper alternative to field training, and as the Services put such systems into place, acquisition managers must add new methods to the traditional technical performance measures to assess the effectiveness of these training systems.

Today each Service has acquisition programs under way to provide distributed-simulation systems for the collective training needs of military organizational units. Because budgets for collective training are tight, one common objective is to maintain or raise unit performance by acquiring comparatively less expensive distributed-simulation training systems to lessen the need for more expensive field training.

Trading field training for distributed-simulation training systems puts pressure on the acquisition community to ensure that the acquired systems are successfully fielded and achieve technical performance objectives. Assessing those objectives may require newly developed measures of performance that mean the same to the acquisition, supporting, and using communities.

The research detailed here examines technical performance measures for distributed-simulation training system acquisitions used for collective training of military units. We discuss the importance of these systems to the acquisition community, using the Army's Close Combat Tactical Trainer (CCTT) as an example. We identify potentially relevant technical performance measures. And finally, we analyze the applicability of the identified technical performance measures during an actual distributed-simulation training system exercise. Findings are generalized to other such systems used for collective training of military organizational units.

THE CLOSE COMBAT TACTICAL TRAINER

Historically, to avoid poor unit performance in combat, military units have focused on training unit tasks through field exercises. With declining budgets, the Services are acquiring distributed-simulation systems that are perceived to be able to train units more cheaply than do field exercises.

All the Services and many joint organizations are acquiring these systems. Under the direction of the Under Secretary of Defense for Acquisition, the Defense Science Board (Foster, 1993) began to define some of the terms applicable to these systems. The Board used the term "distributed" to refer to a "shared battlefield entered from geographically separated sites via communication networks." The Board also defines simulation as a "mix and match of ... simulation methods." Since then the Defense Modeling and Simulation Office (1995) has promoted the simulation "mix and match" concept through "a general purpose architecture for simulation reuse and interoperability" called the high-level architecture (HLA).

"Historically, to avoid poor unit performance in combat, military units have focused on training unit tasks through field exercises."

The Defense Modeling and Simulation Office is developing an HLA that will enable multiple simulation federations (groups of simulations) to exist within and between all the Services, joint commands, and others. Many future distributed-simulation federations and systems are planned for training use (Hammond and Edwards, 1998).

The Advanced Research Projects Agency has composed a joint simulation federation used for collective training—containing air, naval, and army simulated elements within the Synthetic Theater of War initiative (Meier, 1999). Another system, the "distributed mission trainer" (DMT), is a priority for the Air Combat Command (Hawley, 1998). When fielded in 1999, DMT will add integrated and distributed manner simulator systems at Eglin, Langley, Shaw, and Tinker Air Force Bases (AFBs) to simulation systems already in the Air Force (Kuhn, 1998). The integration will provide a complete spectrum of aircraft and facilities for Air Force unit training and mission rehearsal.

An illustrative example of a distributed-simulation training system acquisition is the Army's CCTT, currently being fielded. As do field training exercises, the CCTT "will train Armor, Cavalry, and Mechanized Infantry Platoons through Battalion/Task Force on their doctrinal Mission Training Plan collective tasks" (Hammond and Edwards, 1998).

But unlike field training exercises, "the CCTT-system ... consists of networked vehicle simulator manned-modules ... in combination with Semi-Automated Forces, Combat Support workstations, computer networks and protocols, and After-Action Review systems" (Hammond and Edwards, 1998). Actual military systems like tanks are not used in the CCTT distributed-simulation training system. The CCTT may be considered a distributed, synthetic battlefield with various simulators that enable virtual and other synthetic players to interact in simulated battles.

As in field training exercises, senior evaluators and unit leaders discuss unit task and mission accomplishments and failures with unit members after the CCTT training. This forum is referred to as an "after-action-review." After-action-reviews also provide instruction on process improvements that are aimed at improving overall unit performance.

A simple analogy for this review session might be the discussion that a high school basketball team coach has with his team immediately after a team scrimmage. The emphasis in practice is on processes like individual dribbling and shooting, and team plays like setting up a clear three-point shot or a fast break. The coach does not focus so much on the score (outcome measure) run up against the scrimmage squad, but rather uses those failures and successes as points to correct specific task errors or reinforce successes. In combination with personal and other team tasks and plays, their ability to perform these tasks affects their ability to put points on the scoreboard in the real game.

Unlike most field training exercises, with the exception of some live simulation sites, sophisticated after-action-review systems permit replay of portions of the unit actions that occurred during a CCTT exercise. These after-action review systems enhance unit discussion and further enable unit performance improvements.

As in field training, U.S. Army mechanized and armored units (platoons, companies, and battalions) use mission training plans in the CCTT. These plans identify general and specific tasks with conditions and standards for measuring unit performance against these missions.

Units tend to build ever-higher levels of competence through exposure to ever greater challenges in training (CCTT, 1998). In the dynamics of human and unit growth, the learning environment evolves. From learning basic unit tasks, moving on to learn advanced unit tasks, reinforcement of previously learned tasks, and, finally, integration of various combinations of tasks (typically a mission or set of missions), individuals learn through some combination of instruction, discussion, and exercises.

"These after-action review systems enhance unit discussion and further enable unit performance improvements."

Just as in a field or live simulation exercise, the distributed-simulation training exercise integrates tasks in the form of unit mission scenarios. The training goal is to learn and perfect unit integrated processes like unit tactics, techniques, and procedures that are transferable to many different missions. The focus is typically not exclusively about the resulting outcome for a particular mission. Similar to the basketball scrimmage example, the emphasis is not on the outcome of the scrimmage, but on the processes that can put points on the board during the real game.

Relationship of Systems to the Acquisition Manager

As these distributed-simulation training systems emerge and move toward fielding, the need becomes apparent for metrics to help communicate meaning between dissimilar communities and to evaluate them appropriately. As an example of the importance of metrics,

consider the Williams and Keaton (1998) comprehensive evaluation of the CCTT 1997–1998 initial operational test (IOT) and 1997 limited user test (LUT) conducted by the Test and Experimentation Command (TEXCOM).

During IOT fixed-site simulator training in the CCTT, Williams and Keaton report (based on their aggregate task measures) only a "range of modest to insignificant gains observed during the CCTT training." Specifically, "simulator training during the third through seventh weeks of the IOT indicates that few performance gains were achieved by the units undergoing training."

Despite this recorded lack of performance gains in the CCTT, Williams and Keaton report that "At the aggregate level across all subtasks, the CCTT [-trained] units performed significantly better at NTC" (National Training Center field exercise) than other baseline task forces (Figure 1) (Williams and Keaton, 1998). Specifically, aggregating company team performance within each observed task force, "Task Force 4 [TF4], the CCTT test unit, clearly outperformed the three baseline task forces."

There may be many different explanations for these starkly different observations of performance. One would hope that CCTT training was the primary contributor to success at the NTC. But Williams and Keaton conclude, "The IOT in-simulator performance data was insufficient to demonstrate a linkage between CCTT training and performance attained in the field."

One alternative conclusion is that additional measures and measurement instruments, or a different approach, are needed to capture unit performance improvement that may have actually

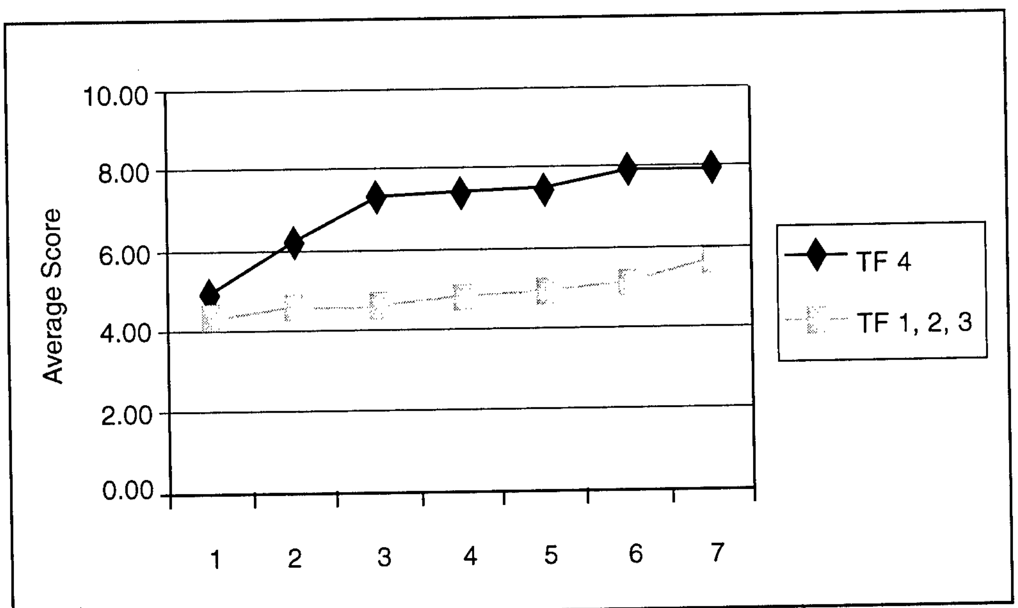


Figure 1. Pure CCTT Company Teams (TF4) versus All Baseline Company Teams (TF 1,2,3)

occurred while training in distributed-simulation systems. Supporting this alternative conclusion is the fact that these systems focus on improving the processes that make a unit function and not so much on the outcome of a particular training scenario.

Since distributed-simulation training systems are acquisitions, both the acquisition community and the training community could benefit from a shared lexicon of technical performance measures to provide a clearer indication of overall potential to achieve ultimate technical objectives.

In 1998 Maj Kenneth Delano published a survey of program managers showing that meeting "technical performance objectives" is ranked first as an indicator of program success. His survey also revealed that program managers ranked their own "ability to communicate" as the most important factor in program success (Delano, 1998). (The article did not define either of these terms more definitively.)

Aiding both evaluation and communications, *The Systems Acquisition Manager's Guide for the Use of Models and Simulations,* published by the Defense Systems Management College (Piplani, Mercer, and Roop, 1994), identifies numerous outcome-oriented, technical performance measures for use by acquisition managers of combat systems.

By contrast, reports on technical performance measures for unit collective training systems are scant if at all present in the acquisition literature. Further, technical performance measures for evaluation of individual training systems have traditionally been submerged within the related combat system acquisition. Typically, training systems were justified as trainers for a specific aircraft, weapon system, etc. Consequently, technical performance measures for individual and crew-training systems have been system specific and oriented to system performance.

TEAMWORK AND TASK PERFORMANCE

The most applicable traditional technical measure identified in the Piplani et al. (1994) publication is an aggregate outcome measure referred to as loss exchange ratio (LER). The LER can be used to judge individual or crew performance improvements. The LER is an outcome measure that compares enemy losses to friendly losses. Using an air warfare analogy, a loss exchange ratio might compare the number of enemy aircraft shot down to the number of friendly aircraft shot down. The more enemy aircraft shot down for every friendly aircraft shot down, the better your system. A difficulty in this approach is that it is limited in scope to comparative systems/units and, in a peacetime environment without actual adversaries, the LER becomes suspect.

"...reports on technical performance measures for unit collective training systems are scant if at all present in the acquisition literature.."

Further, Johnston, Smith-Jentsch, and Cannon-Bowers (1997), Smith-Jentsch, Johnston, and Payne (in press-a) and Brannick, Prince, Prince, and Salas (1995) indicate that "free play" training exercises

produce inconsistent outcomes in the LER when measuring unit performance change from training period to training period, whereas the alternative to "free play"—a structured exercise—was expensive to build and maintain.

As possible supplemental measures, Glickman et al. (1987), McIntyre and Salas (1995), and others discussed the influence of teamwork—a collection of critical behaviors and interpersonal skills—on unit or collective task performance. These two technical performance measures—teamwork and unit task performance—are not widely discussed in the system acquisition literature. As measures they represent analysis of the process as opposed to the aggregate outcome of those processes.

"The use of both process measures to supplement outcome measures allows for a more complete assessment of system contribution...."

Johnston et al. (1997) refined these teamwork dimensions and in a second review, Smith-Jentsch et al. (in press-a and -b) refined the four teamwork dimensions—discussed below—into more reliable and independent dimensions containing sets of specific interpersonal behaviors. Qualitative assessment for each dimension and behavior can be done using behaviorally anchored rating scales (Johnston et al., 1997).

The use of both process measures to supplement outcome measures allows for a more complete assessment of system contribution (Brannick et al., 1995; Johnston et al., 1997; Smith-Jentsch et al., in press-a; Stout, Cannon-Bowers, and Salas, 1997). This is significantly different from the more familiar and traditional emphasis on outcome measures identified by Piplani et al. (1994).

Collective Training in a Distributed-Simulation Training System

So what research could better illuminate the contribution of process measures as a supplement to outcome measures with respect to the discussion and evaluation of distributed simulations used for training?

During our research we observed collective training in the CCTT distributed-simulation training system to gain insight into what teamwork and task performance measures might provide.

Johnson and Noble (1994) indicate that distributed interactive simulation has the potential to effectively train the following primary tasks: command, control and communications (C3); maneuver and navigation; teamwork, and leadership.

For this study, the research team investigated measures for two of these tasks—teamwork and C3 task performance—by observing the normal training of two active-duty U.S. Army battalion task forces within the CCTT facility at Fort Hood, TX. Each battalion task force reported to the CCTT facility to conduct training. The battalion task force received familiarization training on the CCTT and then practiced operating and maneuvering manned module vehicles and units within the CCTT.

In the recorded training exercise, the task force received a "movement to contact" mission and entered its tactical operations planning process. A tactical

plan was devised, rehearsed, and then executed in the CCTT.

Next, each battalion task force conducted an after-action-review feedback session on unit tactical performance at both the company level and the battalion level. After that, the units repeated the movement-to-contact mission. Upon completion of the second trial, another feedback session was conducted to assess tactical performance.

Participants in the study served in duty positions that included company commander, executive officer, and platoon leaders. For the purpose of this study, this team of leaders was referred to as the tactical command and control team, since these individuals provide the leadership to command and control their units while executing their mission.

For control purposes the same movement-to-contact scenario was used between the first and second simulation run. The selected tasks to be performed were identical between simulation runs. The scenario in each run presented the same mission, enemy force, terrain, time frame, environmental conditions, and semiautomated entities' coded behavior. The opposing force consisted entirely of semiautomated forces under the control of an experienced operator.

The semiautomated-force operators used their "free play" prerogative in the second run. Specifically, units typically train against a lesser able opposing force in their initial training. In accordance with the learning objectives of the commanding officer, the semiautomated-force operators typically increase the degree of difficulty by increasing the quality of semiautomated-force tactical operations in subsequent runs (CCTT, 1998).

As indicated above, this common training approach with increasing difficulty in subsequent training exercises was suspected to influence LER relationships. Since the research was aimed at supplementing the LER as a technical measure, the research team collected LER data.

We wanted to evaluate the use of process measures in light of aggregate outcome measures. To facilitate the evaluation, we used an event-based approach to focus on teamwork dimensions and unit task performance of each company's tactical command and control team during each movement-to-contact mission.

"Participants in the study served in duty positions that included company commander, executive officer, and platoon leaders."

Each event contains a unique tactical situation that requires team members to coordinate and exchange information at each step in order to assess the situation, make the appropriate decisions, and execute the correct actions. We selected three specific events that were likely to require the execution of team behaviors. Hence the mission was broken up into three events. The events selected were:

- perform tactical movement (17-2-0301);
- perform actions on contact (17-2-0304); and
- perform an attack by fire (71-2-0311) (Department of the Army, 1988).

Each event has task steps (processes) or subtasks, some which have critical subtasks associated with them. Success or failure of each task step was recorded.

A set of teamwork observation forms (measurement instruments) adapted from methodology used by both Johnston et al. (1997) and Smith-Jentsch et al. (1996a) were applied for each of four teamwork dimensions: communication, information exchange, team initiative and leadership, and supporting behaviors (team initiative/ leadership, Figure 2).

Team behaviors were categorized into effective and ineffective team behavior. Team behavior quality ratings were assessed three ways: the ratio of effective to ineffective team behaviors, the impact or severity of specific team behaviors, and an overall subject-matter expert rating of team behaviors. Quality ratings assessed the four teamwork dimensions and specific team behaviors (Table 1) using a 1 to 5 Likert scale.

Team Initiative/Leadership

Event One. Tactical movement en route to enemy contact: Event One begins at start of scenario and lasts until contact is made with an enemy force.

Task: PERFORM Tactical Movement (17-2-0301) Ref: FM 71-1

Unit: ________________ Simulation Run: ________________

Clear and appropriate guidance provided to team when needed.

1 2 3 4 5

Guidance is unclear or never stated. — Clear and appropriate guidance always stated.

Clear and appropriate guidance provided to team when needed.

1 2 3 4 5

Priorites unclear or never stated. — Clear and appropriate priorities always stated.

Clear and appropriate guidance provided to team when needed.

1 2 3 4 5

Reform inappropriate or never stated. — Reform appropriate and always stated.

Remarks:

Team Initiative/Leadership Frequency

- Provide guidance or suggestions (Effective) 0 1–5 6–10 >10___
- Provide guidance or suggestion (Ineffective) 0 1–5 6–10 >10___
- States clear team/ individual priorities (Effective) 0 1–5 6–10 >10___
- States priorities (Ineffective) 0 1–5 6–10 >10___
- Refocus team IAW situation (Appropriate) 0 1–5 6–10 >10___
- Refocus team IAW situation (Inappropriate) 0 1–5 6–10 >10___

Figure 2. Sample Team Observation Worksheet

Table 1. Teamwork Dimensions and Team Behaviors

Teamwork Dimensions	Definitions of Team Behaviors
Information exchange (Effective behaviors, 1–4; ineffective behaviors, 5–7)	1. Seeks information from available sources 2. Passes information to the appropriate persons 3. Provides accurate "big picture" situation update 4. Accurately informs higher commander 5. Has to be asked for information 6. Provides inaccurate situation update 7. Inaccurately informs higher commander
Communication (Effective behaviors, 1–3; ineffective behaviors, 4–7)	1. Uses proper phraseology 2. Provides complete reports 3. Adequate brevity; avoids excess chatter 4. Uses improper phraseology 5. Provides incomplete reports 6. Uses excessive chatter 7. Communications are inaudible or garbled
Team initiative/leadership (Effective behaviors, 1–3; ineffective behaviors, 4–6)	1. Provides effective guidance or suggestions to team members 2. States clear team and individual priorities 3. Appropriately refocuses team in accordance with situation 4. Provides ineffective or unclear guidance or suggestions to team members 5. States ineffective or unclear team and individual priorities 6. Inappropriately refocuses team in accordance with situation
Supporting behavior (Effective behaviors, 1–4; ineffective behaviors, 5–6)	1. Corrects team errors 2. Requests backup or assistance when needed 3. Provides backup or assistance when needed 4. Provides constructive feedback 5. Fails to correct team errors 6. Provides or uses nonconstructive feedback

In order to avoid inconsistency of assessment between multiple observers, one evaluator was trained and validated at 100% proficiency in identification and classification of teamwork dimensions and respective behaviors by using the "team dimensional training" computer-based-instructional software program (Smith-Jentsch, Zeisig, Acton, and McPherson, in press-b). The same observer assessed team behavior quality ratings and team task performance for all teams.

Results and Analysis

After training was completed, the CCTT after-action-review tapes were analyzed to observe, categorize, and record

observations of teamwork and task performance. Eight company-level movement-to-contact mission scenarios were evaluated. All radio communications and team or unit actions were observed and monitored separately for each tactical command and control team.

Task performance was assessed for each task event based on the *U.S. Army Mission Training Plan for Tank and Mechanized Infantry Company and Company Team* mentioned earlier. Teamwork dimensions and team behaviors were analyzed for indication of improvement. In addition, the traditional loss exchange ratio measure was evaluated.

As a means of analyzing C3 task performance, a series of matched pairs, one-tailed *t*-tests compared the difference in critical task and subtask success between simulation run No. 1 and run No. 2. Matched pairs, one-tailed *t*-tests compared loss exchange ratios differences between runs but due to the nature of the selected tasks, not all tasks involved a LER. For all statistical tests a significant difference was declared if the probability of random occurrence was less than or equal to 0.05.

C3 Task Performance and LER Results and Analysis

C3 task performance indicated statistically significant improvement in mission training plan critical subtask (ρ = .044) and total subtask success (ρ = .007) as shown in Figure 3. For the LER, from a sample of eight tasks that did involve the LER, five teams had an increase in the LER, two teams had a decrease in the LER, and one team had no change in the LER. Statistically the LER did not indicate any change due to the variability in the sample, although the change was relatively large as Figure 3 shows (ρ = .57).

Variations in task difficulty during the second run, due to the free play in the

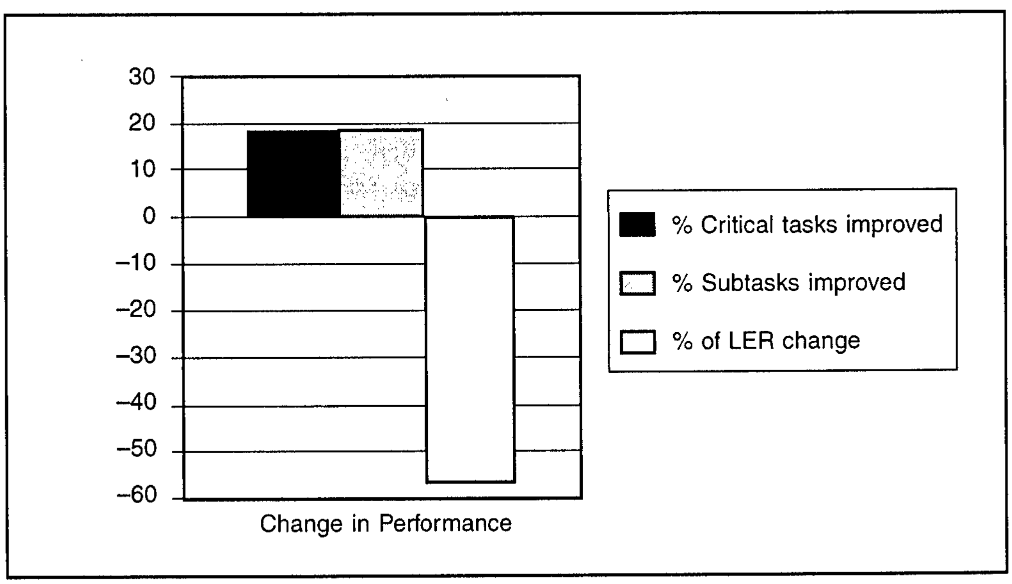


Figure 3. Performance Change between Run No. 1 and Run No. 2 based on Cited Technical Performance Measure

simulation, altered some tactical issues (holding of key [advantageous] terrain, enemy initiating contact from a hasty defense/attack-by-fire positions, force ratio of attacking to defending units, etc.). As suspected, these variations in task difficulty between the two runs may have influenced or confounded the outcomes of the LER data.

Teamwork Quality Ratings

In order to determine if training in virtual simulation resulted in an improvement in teamwork, teamwork quality ratings were assessed for each run. Quality ratings between the first and second runs were found to have improved to a statistically significant degree for all teamwork dimensions. Additionally, improvements in quality ratings for 13 team behaviors that make up the teamwork dimensions were found to be significant.

Conclusions

Whether in athletic competition or in combat, quality teamwork demonstrates its tremendous value. An example process or task worked on by a highly skilled team might be the "no-look pass" between basketball players Michael Jordan and Scottie Pippen. Their teamwork created many national championships.

In the past the U.S. Armed Forces have had significant opportunity to develop expertise in unit teamwork and mission task performance during training involving field operations using actual equipment and formations. That environment has changed significantly due to a number of factors. But despite the change, we don't want to become the Chicago Bulls of 1999. To compensate, our armed forces appear ready to acquire less costly distributed-simulation training systems in order to help fill the gap created by reduced field training.

These findings indicate that training in distributed-simulation systems can result in statistically significant improvements in teamwork, C3 task performance, and, potentially, the loss exchange ratio. Specifically, our research indicates that distributed-simulation training systems can help fill at least two gaps—teamwork and C3 task performance. Statistically significant improvements in the quality of teamwork were shown while conducting training in a distributed-simulation training system. Also, C3 task performance was found to significantly improve between training sessions as shown by increased mission training plan critical subtask and total subtask successes.

Our study also indicates traditional measures such as loss exchange ratios do not appear to be appropriate as sole technical measure when evaluating the suitability of simulation systems used for training. We observed no overall statistically significant change in task performance between simulation runs as measured by the LER. As often the case in training, the LER may not be a credible indicator of improved proficiency of the unit as the difficulty of the opposing force fight might increase for training purposes from run No. 1 to run No. 2. While duplication of the same scenario and difficulty level

"Whether in athletic competition or in combat, quality teamwork demonstrates its tremendous value."

is possible in training, this typically only occurs for units that fail that level and need to be retrained.

The findings provide initial indication that process-oriented measures such as teamwork and subtask and step performance are viable and should supplement the acquisition managers' set of technical performance measures for distributed-simulation training systems. These measures provide the acquisition system manager a more complete assessment of the ability of a prospective distributed-simulation training system than loss exchange ratio would by itself. Further, these measures are intuitive and simple, helping to satisfy the challenge of communications as well as evaluation.

We identify our measures, approach, and measurement instruments, which may prove useful for more general application to other distributed-simulation acquisition involving collective training. Further, they appear appropriate not only for U.S. Army acquisitions but also for the Air Force, Navy, and Joint organizations in light of DMT and HLA efforts. These findings imply that these process-oriented technical performance measures and methodologies may be additional tools with which astute acquisition manager should be familiar.

Further research is required to determine if these findings can be confirmed with larger sample sizes, perhaps over time and across other distributed-simulation systems used for collective training. Further research may address the application of these findings to other training audiences within distributed-simulation such as air wings, ship command and control, higher staffs, and other types of organizations.

Major Michael J. Lipinski currently serves as a Battalion Executive Officer for 1-36 Infantry, Friedburg, Germany. Major Lipinski was commissioned in the Infantry through OCS and is a recent graduate of the Command and General Staff College at Ft Leavenworth, KS. He was also recently selected for advanced civil schooling and graduated from the University of Central Florida with a Masters in Interactive Simulation and Training Systems.

Dr. Michael D. Proctor currently serves as an assistant professor with the Industrial Engineering and Management Systems Department of the University of Central Florida. His research interests include simulation-based acquisition, systems engineering, team training, analysis, artificial intelligence, and test and evaluation. Dr. Proctor is a 1972 graduate of the U.S. Military Academy and retired as a lieutenant colonel in 1995.

(E-mail address: proctor@iems.engr.ucf.edu)

References

Brannick, M., Prince, A., Prince, C., & Salas, E. (1995). The measurement of team process. *Human Factors, 37,* 641–651.

Close combat tactical trainer (CCTT). (1998). Interservice/industry training, simulation and education conference demonstration and presentation. (Nov 30 to Dec 3). Orlando, FL. Proceedings available from National Training Systems Association, Arlington, VA.

Delano, K. J. (1998). Identifying factors that contribute to program success. *Acquisition Review Quarterly, 5*(1), 35–49.

Department of the Army. (1993). *Mission training plan for the tank and mechanized infantry company and company team* (ARTEP 71-1-MTP). Washington, DC: Author.

Foster, J. S. (1993). *Report of the Defense Science Board Task Force on Simulation, Readiness and Prototyping.* Washington, DC: Office of the Under Secretary of Defense for Acquisition.

Glickman, A., Zimmer, S., Montero, R., Guerette, P., Campbell, W., Morgan, B., & Salas, E. (1987). *The evolution of teamwork skills: An empirical assessment with implications for training* (ONR Tech. Report 87-016). Orlando, FL: Naval Training Systems Center, Human Factors Division.

Hammond, A. R., & Edwards, M. (1998). Close combat tactical trainer. http: // www.stricom.army.mil/STRICOM/PM-CATT/cctt.html.

Hawley, R. E. (1998). Keynote address, Interservice Industry Training, Simulation and Education Conference.

Johnston, J., Smith-Jentsch, K., & Cannon-Bowers, J. (1997). Performance measurement tools for enhancing team decision-making training. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications.* Hillsdale, NJ: LEA.

Kuhn, T. Super-duper simulators. *Airman,* http://www.af.mil/news/airman/1198/road.htm.

McIntyre, R., & Salas, E. (1995). Team performance in complex environments: What we have learned so far. In R. Guzzo & E. Salas (Eds.), *Team effectiveness and decision-making in organizations* (pp. 9–45). San Francisco: Jossey-Bass.

Meier, H. (1999). *Lessons learned from synthetic theater of war (STOW 98) and cooperative yearly operation technology evaluation (COYOTE 98) using the DiSECT distributed exercise manager (DEM).* 1999 Spring Simulation Interoperability Workshop, Simulation Interoperability Standards Organization.

Piplani, L. K., Mercer, J. G., & Roop, R. O. (1994). *Systems acquisition manager's guide for the use of models and simulations.* Fort Belvoir, VA: Defense Systems Management College Press.

Smith-Jentsch, K., Johnston, J., & Payne, S. (in press-a). Measuring team-related expertise in complex environments. J. A. Cannon-Bowers & E. Salas (Eds.), *Decision making under stress: Implications for individual and team training.* Washington, DC: American Psychological Association.

Smith-Jentsch, K., Zeisig, R., Acton, B., & McPherson, J. (in press-b). Team dimensional training. In J. A. Cannon-Bowers & E. Salas (Eds.), *Decision making under stress: Implications for individual and team training.* Washington, DC: American Psychological Association.

Stout, R., Cannon-Bowers, J., & Salas, E. (1997). The role of shared mental models in developing team situational awareness: Implications for training. *Training Research Journal, 2,* 85–116.

Williams, D., & Keaton, M. L., (1998). *System evaluation report (SER) of the close combat tactical trainer (CCTT) initial operational test and evaluation.* Alexandria, VA: U.S. Army Operational Test And Evaluation Command.

BASELINING ACQUISITION REFORM

Raymond W. Reig

Accumulating baseline data on the defense acquisition system is essential to gauging just how successful reform efforts have been so far. This article delineates the first step in that process.

Baselining, in and of itself, or as a step in continuous process improvement, has become an accepted modern management technique. Baselining attempts to describe and capture the level of success of an existing system. Then the proposed system changes are applied to the existing system. The changed system should show a large enough increase in success over the existing system to warrant the cost and other expenses of implementation.

The Department of Defense (DoD) Acquisition Reform (AR) Program is a series of changes being incorporated into the DoD acquisition system. This article is an *ex post facto* attempt to baseline the DoD acquisition system prior to the introduction of reforms. In order to do that, we have to determine the effective date of the changes and the level of success of the then-existing DoD acquisition system. Data is available to allow us to do just that.

The only point of this article is to identify and describe the first step in a three-step process called variously benchmarking, baselining, or part of continuous process improvement:

- **Step 1** is to identify a process, procedure, or product into which a series of changes or improvements are to be incorporated. Describe the current process, procedure or product as carefully as possible as regards its current effectiveness and efficiency. Establish a date for this baseline of the existing system.

- **Step 2** is to introduce the changes or improvements into the process, procedure or product.

- **Step 3** is to measure the effectiveness and efficiency of the changed system, at some future date.

Discussion

First we shall establish the effective date of acquisition reform. Table 1 chronologically lists most of the AR policy guidance and other major events of the program. It is important to note that we are seeking the first date that AR policies could be considered effective in the field; that is, the approximate date that AR initiatives began being implemented in a significant number of program management offices, and other field acquisition organizations. In the January/February 1997 *Program Manager* article, Doreen Harwood states that "A gap of as much as six months can occur between the time a statute or policy change is issued before it is received in the field" (Harwood, 1997, p. 41).

Colleen Preston, who was the designated change agent to direct the acquisition reform program, was appointed Deputy Under Secretary of Defense for Acquisition Reform (DUSD[AR]) in June 1993. This could be considered the start of acquisition reform, but certainly not the date that it first became effective in the field. Preston outlined the initial efforts of acquisition reform four years later (1997, pp. 25–26). I have inserted the dates of each action mentioned (Table 1).

> We started off initially with following up on the Section 800 Panel recommendations [January 1993]... So we made that the initial thrust. For the first year we practically did nothing but focus on that legislative effort day-to-day...Then we started working the Process Action Team (PAT)...We started with electronic commerce because that was critical... Then...Specification and Standards issue [June 1994] we took on as our second PAT...Then we looked at...contract administration, the procurement process and...oversight and review of the systems acquisition process [December 1994]. That particular PAT process was very difficult because it focused on the relationship between OSD and the Services.

Other milestones include the May 10, 1995, memorandum of William Perry, then Secretary of Defense, implementing the integrated product team (IPT) concept within DoD; the initiation of cost as an independent variable (CAIV) in December 1995; and the release of the new 5000 Series acquisition documents in December of 1996. Several other Milestone dates can be extracted from the chronology, but these show the time required for approaching effective AR implementation in the field. An alphabetical list of AR initiatives is in Figure 1.

Throughout the AR implementation period there was recognition that for this "cultural change" to be effective there had to be visible and continuous support from the top and available tools to understand what was desired. Again, the chronology shows us the many initiatives taken along these lines, such as: Paul Kaminski hosting a one-day DoD offsite ("Institutionalizing IPTs") on July 20, 1995;

Table 1. Chronology of Acquisition Reform Events

Date	Event
February 1991	DoDD 5000.1 DoDI 5000.2 changed and reissued and 5000.2M promulgated.
January 1993	The Acquisition Law Advisory Panel (Section 800 Panel) findings reported to Congress.[a]
June 1993	Colleen Preston assumes the position as Deputy Under Secretary of Defense for Acquisition Reform.[a]
October 1993	Federal Acquisition Streamlining Act (FASA) of 1994 enacted.[a]
First quarter 1994	The Advanced Concept Technology Demonstration program initiated.[b]
February 1994	William J. Perry replaces Les Aspin as Secretary of Defense.[a]
February 1994	Secretary Perry issues "Acquisition Reform, A Mandate for Change."[a]
March 1994	Secretary Perry attaches "Mandate for Change" to a letter to the leadership of the Department of Defense.[a]
June 1994	Preston authors an article, "Acquisition Reform—Making it a Reality," in *Phalanx: the Bulletin of Military Operations Research* (June 1994, 27[2]). The article concludes with a section titled, "How Can You Participate?"[a]
June 1994	Secretary Perry issues memo: "Specifications and Standards—A New Way of Doing Business."[a]
October 1994	Paul Kaminski sworn in as Under Secretary of Defense for Acquisition and Technology (USD[A&T]).[a]
c. 1994	DUSD(AR) position to report to USD(A&T).[a]
December 1994	The Oversight and Review of the Systems Acquisition Process PAT report published.[b]
December 1994	The Defense Acquisition Pilot Program launched as allowed by FASA.
March 1995	USD(A&T) establishes an IPT for the purpose of rewriting the February 23, 1991, 5000 Series documents.[b]
April 1995	Kaminski issues a memorandum, "Reengineering the Acquisition Oversight and Review Process." First recommendations of the PAT team approved.[b] (continued)

Date	Event
May 1995	Secretary Perry implements the IPT concept for DoD via a memorandum.[a]
July 1995	Kaminski holds a DoD offsite entitled "Institutionalizing IPTs—DoD's Commitment to Change."[b]
November 1995	Rules of the Road: A Guide for Leading Successful Integrated Product Teams is published.[b]
December 1995	CAIV was initiated.[a]
December 1995	USD(A&T) issues guidance for making "class action" contract changes to existing contracts on a facility-wide basis. AKA Single Process Initiative (SPI).[b]
February 1996	*DoD Guide to Integrated Product and Process Development,* (Version 1.0) issued by the OUSD(A&T).[a]
February 1996	Director, Test, Systems Engineering, and Evaluation, publishes DoD Guide to IPPD, Version 1.0.[b]
March 1996	Update of the DoD 5000 Documents approved by the USD(A&T), DOT&E, and ASD (C3I).[a]
March 1996	The ODUSD(AR) produces the video The Overarching and Working Level Integrated Product Teams, and the OIPT-WIPT Information Guide.[b]
April 1996	DoD and Texas Instruments sign first SPI agreement for manufacturing standards for all its products.[b]
May 1996	DoD Acquisition Reform Day is held.[b]
July 1996	The *Defense Acquisition Deskbook,* first piece, released.[b]
September 1996	Kaminski's memorandum provides guidance for dealing with specification or process changes on subcontracts (SPI).[b]
December 1996	The publishing of DoD 5000.2R, *Mandatory Procedures for Major Defense Acquisition Programs (MDAPs) and Major Automated Information Systems (MAIS) Acquisition Programs.* (Includes change 1).[a]

[a] All data and information obtained from Defense Systems Management College. (1997, December). A Model for Leading Change: Making Acquisition Reform Work (report of the 1996-1997 DSMC Military Research Fellows). Fort Belvoir, VA: Author.

[b] All data and information obtained from Defense Systems Management College. (1997, January-February). Acquisition reform —the end of the beginning. *Program Manager* (special issue), 26(1).

Acquisition Law Advisory Panel (Section 800 Panel)
Audit/Inspection Reform
Advanced Concept Technology Demonstration (ACTD)

Buying Commercial

Defense Acquisition Deskbook
Defense Acquisition Pilot Programs (DAPP)
Direct Vendor Delivery
Dual-Use Technology

Electronic Commerce
Empowerment

Federal Acquisition Streamlining Act of 1994 (FASA)

Integrated Product and Process Teams

Lean Logistics

Modeling and Simulation
Multiyear Contracts

Outsourcing

Partnerships with Industry
Performance-Based Contracting
Privatization
Program Stability

Reduced Government Oversight
Reengineering the Acquisition Oversight and Review Process

Single Process Initiative (SPI)
Specifications and Standards Policy
Streamlined Solicitation Packages

Update and Reissue of the DoD 5000 Series Documents

Workforce Education

Figure 1. Alphabetical List of Acquisition Reform Initiatives

publishing *Rules of the Road: A Guide for Leading Successful Integrated Product Teams* in November 1995; the DoD AR Day of May 31, 1996; and the release of the *Defense Acquisition Deskbook* in July 1996.

From all of this it is possible to estimate when AR could first have had a practical effect in the field. For me, this date is on or about January 1996, although efforts continue beyond this date, and will into the future. January 1996 is the

estimated date that a sizable number of the acquisition workforce in the field could first be expected to put acquisition reform into practice. The date we are more interested in, however, is the first probable date that AR could affect a major defense acquisition program.

Prior research has shown that the average length of an Acquisition Category I (ACAT I) program in the engineering and manufacturing development (EMD) phase is 7.4 years (Gailey, Reig, and Weber, 1995). The Milestone III (MS III) decision generally concludes the EMD phase, where design and cost-impact decisions determine more than 80 percent of the total life-cycle system costs. If a program's MS III is 6 months away, there is little chance AR changes will affect that program in EMD. Therefore, if acquisition reform first became effective within the field acquisition workforce on or about January 1996, it could only have an EMD effect on programs whose MS III Defense Acquisition Board (DAB) meeting was after July 1996. The MS III program date is important because that is the point at which we will measure the success of the then-existing DoD acquisition system.

"In 1991, two significant events occurred that, in retrospect, embodied the "cultural change" so central to acquisition reform."

The influence of two new initiatives on the effective date of AR for any particular program has not been discussed. These initiatives are the Advanced Concept Technology Demonstration Program (ACTD), and the Defense Acquisition Pilot Programs (DAPPs), both initiated in 1994. Ten ACTDs were initiated in fiscal year 1995, and 12 in fiscal year 1996. This is a very small number of programs when compared to the approximately 200 programs listed at any one time on the Office of the Secretary of Defense Test and Evaluation (OSD T&E) program oversight list. Extracts of a report on the lessons learned and recommendations on how to proceed with DoD-wide implementation of the DAPP initiative are shown in Appendix A. An exhaustive review of the effects of DAPP programs on acquisition efficiency is beyond the scope of this research, however. Other ongoing Office of the Secretary of Defense (OSD) review and analysis efforts explore this area in much greater depth.

Although not considered part of the acquisition reform program, the first event listed in Table 1 is worthy of discussion. In 1991, two significant events occurred that, in retrospect, embodied the "cultural change" so central to acquisition reform. The first of these, the February 23, 1991, revision to the 5000 Series documentation, modified an earlier version of the 5000 Series that required important user requirements like the operational requirements document (ORD) and initial operational capability (IOC), to be stated at MS I. Prior to February 1991, it was generally understood that these requirements were firm and not subject to change. In the revision, the new 5000 Series stated these and other requirements were subject to review and change if necessary at each Milestone. This allowed for a more reasoned approach to changing requirements as more data accumulated, and allowed the program manager to suggest changes in a more receptive environment. Changing user requirements if necessary

at stages in the development of a system became similar to the latter-day CAIV approach, albeit for a less focused reason.

Also in 1991, the commanders of the Service operational test activities (OTAs) realized that they no longer could operate in the mode of being the independent director of the "final exam," the initial operational test and evaluation and the operational evaluation (IOT&E/OPEVAL) just prior to MS III. Rather, they initiated an earlier consultative role with the developing activity and, by means of early operational assessments, worked with the program managers to clarify what would be expected at the IOT&E/OPEVAL. This change in *modus operandi* occurred before the AR initiative of IPTs, but clearly achieves some of the same desirable objectives.

Because of these two activities, the introduction of a "forerunner CAIV" and a "forerunner IPT," the year 1991 could be dubbed the year of the unheralded AR. Because this was a revision to the established 5000 Series system, the lead time to be effective in the field might be half that required for the later, entirely new, more expansive AR changes.

With the above rationale, the first programs that would be beneficially affected by AR would be those whose MS III DAB occurred in July 1996 or afterward. Therefore, the baseline of interest would be the success level of programs whose MS III DAB was prior to July 1996. This was the success level of the DoD acquisition system prior to AR changes. To measure program success, we will use the standard parameters of cost, schedule, and performance of a program during EMD. Both cost and schedule successes are obtained by a review of a program's selected acquisition reports (SARs) during EMD, and performance success will be based on a review of the program's IOT&E/OPEVAL test reports and the associated beyond low-rate initial production (BLRIP) evaluation issued by the Director, Operational Test and Evaluation (DOT&E).

Fortunately, these data have been obtained for a different purpose over the past four years. A technical report has been published detailing the concept and rationale behind the data obtained to date (Gailey, Reig, and Weber, 1995, pp. 3.2–3.3). For each of the parameters—cost, schedule, and performance—we have assigned a success rating of from 5 (very good) to 1 (poor). The cost and schedule data from the Blue Books (no longer maintained) or the SARs are objective, using the DoD standard decrements of 15 percent in cost and 6 months in schedule. Performance success ratings were subjectively assigned using descriptive criteria that delineated between the five possible ratings. Operational test reports addressed operational effectiveness and operational suitability. For our research purposes, we have assigned a third rating, overall operational success, which results in three performance success rating from the Service OTA and three from the DOT&E on the OSD staff. In practice, the Service OTA conducts the operational test, but by law, the DOT&E

"To measure program success, we will use the standard parameters of cost, schedule, and performance of a program during engineering and manufacturing development."

must provide an independent evaluation of the test adequacy and the operational effectiveness and suitability of the system under test.

For other purposes, we have gathered EMD cost, schedule, and performance data on programs during the years 1980 to 1996. We are in the process of obtaining the same data for programs as they complete their EMD development phase from 1996 on. The resultant success ratings are shown in Table 2, and the percentage equivalents of the success ratings are shown in Table 3. We have listed two periods from which to choose the performance success level of programs in EMD. The first line in Table 2 has the success ratings for all programs in the data base from 1980 to July 1996. The second line shows the results for the 4 years prior to the date AR could have influenced systems in EMD. The data base allows other, different periods to be used.

CONCLUSIONS

Here we have looked only at the cost, schedule, and performance success ratings of programs whose EMD phase ended prior to July 1996. After this date, AR efforts began to have an effect on major defense acquisition programs. This established the performance level of the DoD (pre-acquisition reform) acquisition system. The improvements in the acquisition system due to ARs can be measured in a few years using a similar methodology. It is probably too early to attempt to measure the improvements in cost, schedule, and performance success of the weapon systems being procured by the DoD acquisition system. But if the research database established at Defense Systems Management College continues to grow, in the future such data will be at hand.

We have not attempted to analyze any of the many other parameters of AR, such as value added due to a better educated workforce, etc. Figure 1 lists all or almost all of the initiatives considered to be part

Table 2.
Average Baseline Success Ratings - Acquisition Reform Baseline

M/S III Time Period and Number of Programs	Cost Success	Cost Percent Overrun	Schedule Success	Schedule Percent Overrun	IOT&E/OPEVAL Results			DOT&E BLRIP Evaluation		
					Effectiveness	Suitability	Overall	Effectiveness	Suitability	Overall
1980-Jul 96 (n=42)	3.6	31%	2.6	64%	4.0	3.7	4.0	3.8	3.6	3.8
Jul 92-Jul 96	3.5	46%	2.1	81%	4.3	3.9	4.2	3.7	3.7	3.8
Post Acquisition Reform Success										
Aug 96-ON	*	*	*	*	*	*	*	*	*	*

* = To be determined

Table 3.
DSMC EMD Research - Success Ratings and Percent Equivalents

Success Rating 5 (very good) – 1 (poor)	Percent (%) Equivalents
5.0	100
4.8	96
4.6	92
4.4	88
4.2	84
4.0	80
3.8	76
3.6	72
3.4	68
3.2	64
3.0	60
2.8	56
2.6	52
2.4	48
2.2	44
2.0	40
1.8	36
1.6	32
1.4	28
1.2	24
1.0	20

of the AR program and may be of some interest in itself. There are other OSD analyses in existence that are much broader in scope and level of effort; these give the reader a broader look at AR improvements.

However, using the results achieved by the pre-AR DoD acquisition system in the 4 years before AR began, and converting the success ratings in Table 2 to percent equivalents (Table 3), schedule success—approximately 42 percent successful in the old DoD acquisition system—could certainly be improved by AR efforts. Cost success at 70 percent should also be amenable to improvement. Future improvement in performance success, where the overall success level was 76–84 percent, may be more challenging.

REFERENCES

Harwood, D. (1997, January-February). Defense Acquisition Deskbook—An acquisiiton reform unqualified success. *Program Manager (special issue)*, *26*(1), 40–42.

Preston, C. (1997, January–February). Acquisition reform. *Program Manager (special issue)*, *26*(1), 22–31.

Gailey, C., Reig, R., & Weber, W. (1995, May). *A study of the relationship between initial production text articles used in a system development program and the success of the program* (DSMC Press Technical Report TR2-95). Fort Belvoir, VA: Defense System Management College.

Raymond W. Reig, B.S., M.I.E., M.P.A., M.B.A., is a visiting research professor at DSMC. He is a graduate of the U.S. Naval Academy, the Air War College, and the DSMC Program Management Course. He was the first Chairman of the Test and Evaluation Department and now works within the College research division. His more than 40 years of experience spans military, government, and private aerospace industry.

(E-mail address: reigr@dsmc.dsm.mil)

APPENDIX A

REVIEW OF DEFENSE ACQUISITION PILOT PROGRAMS

The Defense Acquisition Pilot Programs (DAPP) was initiated as a result of the Federal Acquisition Streamlining Act (FASA) of 1994, which authorized the DoD to experiment with new ways of doing business to achieve the objective of buying systems better, faster, and cheaper. Seven DAPPs are reported on and data are being collected on what aspects of the program are successful, and by how much (metrics). Six of the seven programs are: Commercial Derivative Engine, Defense Personnel Support Center, Hercules (C-130J), Joint Primary Aircraft Training System, Fire Support Combined Arms Tactical Trainer, and Joint Direct Attack Munitions. This exhibit contains my impressions of data whose source is the *1996 Compendium of Pilot Program Reports* issued by the Pilot Program Consulting Group.

The Commercial Derivative Engine (CDE) is produced by Pratt and Whitney and used on the C–17 aircraft. In 1980, McDonnell Douglas Aircraft Company selected the engine to power the C–17, and thus the program was well under way when it was nominated in July 1993 to be a pilot program to "demonstrate the advantages of using derivatives of commercial engines to satisfy military requirements." This program may best exemplify the "buy commercial" initiative of acquisition reform.

The Defense Personnel Support Center (DPSC) began reengineering its business practices in 1989, and the designation as a pilot program facilitated further expansion of these initiatives.

> Continued participation as a regulatory DAPP allows evaluation of statutory relief provided commercial item acquisition by FASA.... The 1995 report highlighted the fact that regulatory relief by itself was insufficient to make a dramatic impact.... It is the acquisition reform environment itself that has contributed the most to the continued success...of these initiatives.

This pilot program may exemplify the difficult-to-measure improvements resulting from a DoD acquisition "cultural change."

The Hercules (C–130J) was designated a DAPP program in September 1995 and "serves as the first major procurement that can draw upon the new commercial practices implemented by FASA." This ACAT II program has developed a comprehensive set of specific and bridge metrics for overall DAPP program guidance. These metrics primarily address business practice issues, but contain two performance metrics. The business practice metrics appears to have been achieved very nicely, but the two performance objectives are to be determined.

The Joint Primary Aircraft Training System (JPATS) is an ACAT IC program whose 9-year engineering management and development phase contract was signed in February 1996. "JPATS specific metrics were developed...to reflect the

unique commercial aspects of the program...and include:

- regulatory and statutory relief,
- program costs,
- RFP preparation and content,
- funding stability,
- ground-based training system,
- earned value reporting system,
- program office staffing,
- contractor team composition, and
- contract administrative services."

Again, these metrics primarily address business aspects. The two performance bridge metrics used are Anthropological Accommodation, and Birdstrike Capability @ 270KTAS.

The Fire Support Combined Arms Tactical Trainer (FSCATT) RFP was designed to incorporate DAPP acquisition reform initiatives.... The contract, fixed-price, 7-year period of performance was awarded in June 1995. "As a DAPP, [the program] is intended to demonstrate that the concepts of dual-use technology could be applied to a defense program in addition to demonstrating the capability of integrating commercial and nondevelopmental item components into a complete system." This is an ACAT III program. The program metrics consist entirely of business practice parameters, with no performance metrics, except for the number of work hours for quality assurance and test and evaluation. Business metric results to date are good, with cost savings estimated at 13.5 percent and a planned schedule reduction of 33 percent.

The Joint Direct Attack Munitions (JDAM) may be the most discussed and closely followed DAPP program to date.

> The nature of JDAM suggests that the major savings...would accrue in production rather than development.... The [program] is intended to demonstrate the prime contractor and key subcontractors are able to develop the JDAM using practices, ...from their commercial sector business base.... Finally, the program is expected to meet the planned development schedule without the process delays that have been incurred on other major defense programs.

JDAM is an ACAT I program whose Phase I EMD effort began in April 1994. Phase II EMD started in October 1995 with the down-select to one contractor. The Phase II (EMD) contract is CPFF with a period of performance of approximately 3 1/2 years. "Thus, the cultural change provided by pilot programs designation had the greatest impact on JDAM and resulted in the implementation of [several new] business practices." The cost metric includes total program costs, but Table 1.3.2 shows Phase II EMD (RDT&E) cost avoidance's to total $49.8M, including $7.3M attributable to reduced wind tunnel tests, $30M of reduced open air test A/C and test units, and $12M due to schedule acceleration.

In discussing program operational performance and cost, the report states

"...large reductions in unit price were possible by trading other 'performance' requirements." Two meaningful performance metrics—end game accuracy and reliability—are being tracked. The JDAM program 1996 status report concludes:

> The most dramatic results of the JDAM acquisition reform efforts are the reported reduction in total program costs of $2.96 billion Then Year (TY) over a 10-year production cycle, a 39 percent reduction in contract administration hours to date, and a 35 percent reduction in development time. Clearly JDAM results to date demonstrate the applicability of commercial practices and other innovative management practices to major defense acquisition programs and the efficiency gains that can be achieved. As reported by the program office, acquisition reform combined with common-sense management is enabling JDAM to realize substantial in-house efficiency gains, reduce contract costs, and improve cycle times.

Conclusions

Have these Defense Acquisition Pilot Programs had an effect on the selection of the cost, schedule, and performance baseline of the pre-AR DoD acquisition system? I believe not, since two of the DAPPs pre-date acquisition reform considerably, three are not ACAT I programs, one contains possibly minor performance metrics, and three contain no performance metrics at all. By contrast, the entire thrust of this article is the combined measurement of cost and schedule data (from SARs), and performance data (from operational testing reports) of ACAT I programs in EMD prior to the effective start date of acquisition reform. A more reasonable comparison would be DAPP program performance as measured by other "standard" programs being developed in the current acquisition reform era.

SYSTEMIC FISCAL OPTIMISM IN DEFENSE PLANNING

Dr. Leland G. Jordan

Defense planning and budgeting increase national security costs by significantly overestimating available future resources. An analysis of Department of Defense out-year resource estimates over a period of 20 years and six administrations—the first econometric analysis of budgeted and realized resources in defense—demonstrates that an optimistic bias has spanned administrations and appears to be a systemic characteristic rather than a political one. The result has significant implications for reduction of defense costs without loss of capability.

Some analysts have suggested that fiscal optimism in defense planning and budgeting results in less defense than could have been achieved given the resources available. That is, fiscal optimism results in less bang for the buck, rather than the more bang for the buck traditionally sought by the Department of Defense (DoD). Programs established under a projected fiscal regime with more resources than later are realized may become unaffordable under the tighter resource constraint. Unaffordable means that the budget is not sufficient to carry out the program at the rate, at the unit cost, and in the quantities originally programmed and planned.

The traditional issue of weapons systems cost growth and the issue of DoD's consistent forecast that it will receive significantly more budgetary resources than it does receive are not separate. The gap between planned and realized budgetary resources is the predominant cause of weapons systems cost growth.

Franklin Spinney addressed the problem of cost growth and fiscal optimism in the early 1980s. His analysis was not well received within DoD, although it achieved sufficient notoriety outside DoD: He was pictured on the cover of *Time* magazine (Isaacson, 1983). He addressed the force structure and unit cost problems that result from optimistic assumptions about the cost progress curves[1] and the reluctance to terminate systems that, although well along in development or production, appear unaf-fordable given the resources actually appropriated (Spinney, 1980). Spinney did not address how

DoD consistently gets into the position of not having enough resources to complete what it has started.

Gansler approaches the issue through the effects on weapons system costs and on strategy and the ability to support strategy (Gansler, 1989, chap. 5). Focusing primarily on management within DoD and on the interface with industry, Gansler addresses "optimistic planning," but does not directly address the source of fiscal over optimism. The Packard Commission identified the problem of optimistic planning and recommended some improvements, but did not present an analysis demonstrating the persistence of the phenomenon across time and administrations (Packard, 1986). Efforts to assess the dollar effect of optimistic planning have been rare and have not been published in the academic press. For example, Rolf Clark's papers, prepared under the auspices of the DoD's Defense Systems Management College and circulated within the DoD, were not published in peer-reviewed journals (Clark, 1990a, 1990b).

This article provides an assessment of the quality of the defense out-year resource forecasts from a system perspective, identifies the source of forecast errors, and draws implications about their costs and the potential for improving the forecasts. Its broader purpose is to identify the nation's out-year budgeting practices as an important area of research in which analysts can contribute significantly to the national welfare. Budgeting, whether for next year or longer periods, is an accountancy function directly affecting management; it should be addressed with the same rigor as is applied to stock price movements, earnings forecasts, and the effects of revised standards.

This analysis is based on the following axioms. If one plans to have significantly more resources than become available then it should not be surprising if the plans are unaffordable. The planner should learn from such experience and begin to estimate better the future resources. We should not expect a perfect forecast, but should expect the quality of the forecasts to improve over time.

Forecast accuracy is especially important for national defense when erroneous forecasts contribute to a lesser capability than could have been obtained at the realized resource level.

The analysis presented here concludes that the defense planning and budgeting system is optimistically biased and that the bias has spanned several administrations. Nonetheless, out-year forecasts have been significantly better under some administrations than under others. Those administrations having demonstrated the greatest bias in their real growth projections also have experienced the greatest shortfalls in resources, implying the greatest impact on management. A proportion of the forecast error can be reduced and improvements (discussed below) can be instituted.

Neither this analysis nor those cited suggest that optimistic planning results from malicious intent. Rather, it is the result of a highly complex system that does not function as intended.

Optimism is defined as a form of the bias discussed in the conceptual statements of the Financial Accounting Standards Board (FASB). Were we able to place a probability distribution on the fiscal projections of the defense budget, we would find that those projections consistently are greater than the expected value.

No attempt is made to assign the causes of the bias to the elements of measurer bias or measurement bias. Measurer bias results when the measurer misapplies the measurement methodology. Such misapplicaion may derive from lack of skill or lack of integrity, or both. Measurement bias results from inadequacy, or lack of validity, of the measurement instrument or method. However much the resultant bias may originate in each of those two causes, it remains a systemic characteristic of the national security planning and budgeting system (FASB, 1985).

THE IMPACT ON MILITARY EFFECTIVENESS

Planning for more resources than become available results in programming a larger force and more investments than can be supported. The defense literature has noted the effects of that discrepancy. Kevin Lewis, in "The Discipline Gap and Other Reasons for Humility and Realism in Defense Planning," concludes that the likelihood of the DoD's planned program achieving its planned effectiveness is small (Lewis, 1994). It is important to recognize that Lewis has in mind the military effectiveness of the forces that result from the plans. Jacques S. Gansler deals with the effects on weapons system costs and on strategy and the ability to support strategy (Gansler, 1989). Spinney also has addressed these effects (Spinney, 1996).

In defense planning, the mix and deployment of forces is optimized within the expected resource constraints.[2] The mix of forces varies as a function of the total financial resources available. For example, a specialized aircraft or other weapons system may be effective and affordable only if it exists in the force in some minimum quantity. Fielding of the system requires development of doctrine and tactics and also the training of the forces and the commanders. In the highly integrated modern battlefield, development and management of compatibility with the associated forces also is required. Clearly, it could be ineffective and cost prohibitive to do all those things for a single aircraft, especially if some backup weapons system were required in the event that single aircraft were lost. At some point, the cost-effectiveness of a specialized system, available in a minimum quantity, is less than the cost effectiveness of the alternative multipurpose weapons system.

"Planning for more resources than become available results in programming a larger force and more investments than can be supported."

Decisions to produce a special-purpose weapons system or the alternative multipurpose system are made on the basis of projected resources. Even once it becomes clear that resource projections were optimistic, reversing such decisions is difficult. The difficulties arise from the added costs incurred by a termination, both economic and psychological, and from the time-lag that would be incurred in developing the multipurpose system. In fact, that time-lag may preclude fielding of the alternative capability soon enough to counter the threat. Thus, the ability to repair a bad decision in response to near-term information about resource availability is limited.

Given the earlier decisions, made on the basis of optimistic resource projections,

the best possible defense program may be significantly less effective than would have been possible had the earlier decisions been made in the context of realistic resource constraints. That situation is modeled below:

Let *E* (year, resources, period) represent the maximum effectiveness of the defense program resulting from decisions made in year *i*, given multi-year projected resource constraint *j*, and serving in the future period *k*. The period may be a specified Future Years Defense Program (FYDP)[3] period or some longer time span (such as a decade). Then, the maximum effectiveness of a defense program, given resource constraint j_1 is

$$E(i, j_1, k) \qquad (1)$$

and the maximum effectiveness of the next year's defense program, covering the same period *k*, but with a revised resource constraint $j_2 < j_1$ is

$$E(i + 1, j_2, k) < E(i, j_1, k). \qquad (2)$$

Some observers have identified the revised planning that results from correction of $E(i, j_1, k)$ to achieve $E(i + 1, j2, k)$ as the source of the acquisition turbulence so roundly condemned by the Packard Commission (1986). Clearly, if the effectiveness decline applies to the next year's program, it also applies to the $i + n$ program where n is an integer greater than one and less than some integer representing the time to develop and field an improved mix of forces.

Because the time to develop and field a weapons system is at least 10 years, the effectiveness decline persists for about that same period.

OTHER ANALYSES OF PLANNING BIAS

The idea that a bias in planning may exist is not new. Henri Thiel (1971) discusses the measurement of such bias and offers several examples of systemic bias. His discussion, because it uses Dutch national forecasts as an illustrative case, establishes the relevance of that technique to the analysis presented here. J. Chapman (1981) applied Thiel's technique to assessment of the accuracy of revenue forecasts by California cities before and after the passage of Proposition 13. He found a tendency toward underestimation of revenues both before and after passage of Proposition 13. Chapman's findings are not directly relevant to this analysis, but his application of Thiel's technique is.

Allusions to bias in national forecasts in the United States are not unusual. For example, J. Sessel (1995) quotes comments by two well-known observers on the White House and Congressional Budget Office (CBO) forecasts. Former CBO Director June O'Neill said, "The history over the past 20 years is that both of us are too optimistic." Alan Auerback, an economist at the University of California, Berkeley, commented, "I've become convinced that there's a pervasive tendency towards overoptimism in both agencies" (Sessel, 1995). In *Affording Defense,* Gansler (1989, chap. 5) refers to "optimistic planning." One of the threads of his analysis is the effect of planning for a greater financial resource than becomes available. Gansler is unusual in his recognition of the adverse effects of such optimism.

The existence of such a systemic bias is relevant to other organizations, both public and private, and knowledge about

the detection and correction of such biases would be an important contribution to the knowledge about managing complex public and private organizations. The magnitude of the effect on other organizations probably is related positively to their planning horizon.

BIAS IN DEFENSE PLANNING

DATA

For this study we examined data for a period of 20 years: fiscal year 1975 through fiscal year 1995.[4] Planned resource levels were compared to the actually available resource levels for the administrations of Presidents Gerald Ford, Jimmy Carter, Ronald Reagan, George Bush, and for President Bill Clinton through fiscal year 1995.

The projected fiscal resources against which plans were constructed consistently exceeded the fiscal resources that actually became available. The situation is portrayed graphically in Figure 1. That figure presents the actual and planned data for President Reagan's second administration, 1985 through 1988.

The bars in the chart show the resource levels for each year of the DoD's planning period. Because a new planning period begins yearly, the bars for each year represent plans from several prior years. The line represents the funding appropriated by the Congress, the fiscal

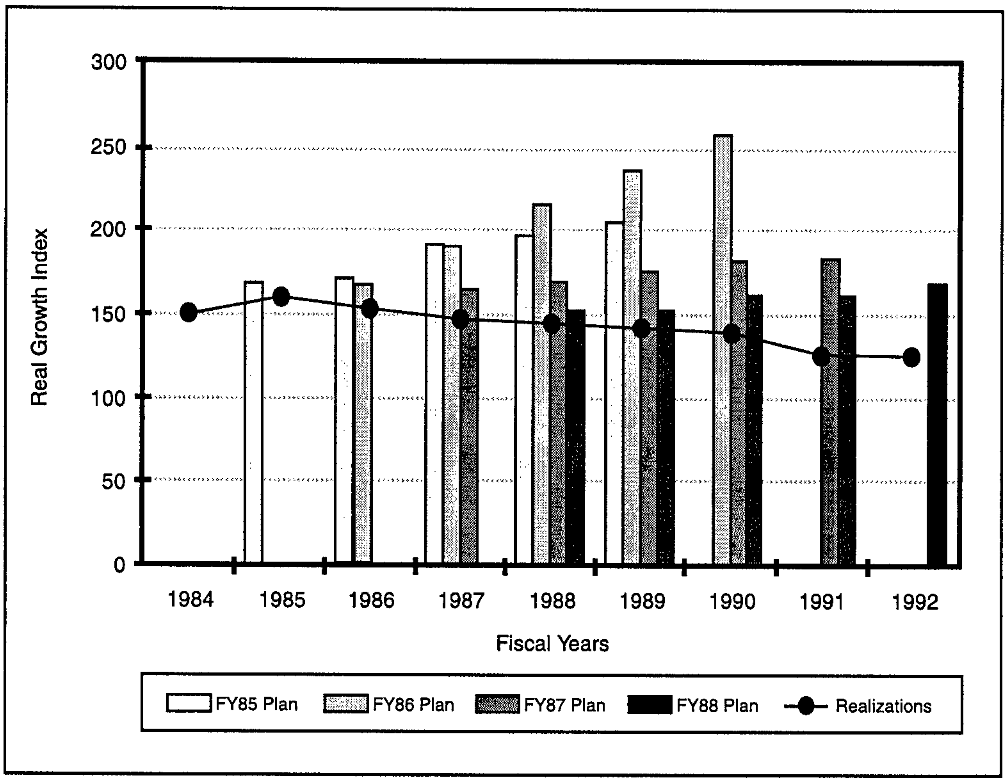


Figure 1. Planned Versus Realized Real Growth, Fiscal Years 1985–88

resource that was realized. Clearly, the plans of each administration extend into the subsequent administrations and, just as clearly, are revised by those subsequent administrations.

In order to remove the effects of inflation, real dollar levels indexed on fiscal year 1974 are plotted. The DoD deflators were applied to the actual appropriations. Those projected at the time of each plan were used to deflate the resource projections, and then were linked to the same deflators that were applied to the series of actual appropriations. Thus, each year's resource levels, projected and actual, were restated in the same dollars and then indexed on fiscal year 1974.

METHODOLOGY

Spinney (1992) used a primarily graphical analysis in his presentations, accompanied by discussion. Figure 1 similarly portrays the data. Gra-phical portrayals provide an intuitive feel for the situation, but they do not support conclusions about the un-derlying causes of the forecasting errors.

In *Applied Economic Forecasting,* Thiel (1971, p. 32) develops a method for analyzing the adequacy of economic forecasts. Thiel decomposes the squared error of the forecast into coefficients related to the sources of the forecasting error.

Our analysis is based on real growth rates, projected and actual, to remove the effects of inflation and also because the projection methodology used in the DoD is based largely on assumptions of future real growth. The analysis uses the natural logarithms of the real growth rates. Their use ensures that the levels in years t_1 and t_2 are the same if the log changes in those years are equal but of opposite sign (Thiel, pp. 47–50).

SOURCES OF THE PROJECTION ERRORS

Thiel's coefficients are derived from the sum of the squared errors as shown below in Figure 2. The coefficients represent bias, variance, and covariance, respectively.

$$\frac{1}{n}\Sigma(P_i - A_i)^2 = (\overline{P} - \overline{A})^2 + (S_p - S_a)^2 + 2\ 1 - r) S_p S_a \qquad (3)$$

$$1 = \frac{(\overline{P} - \overline{A})^2}{\frac{1}{n}\Sigma(P_i - A_i)^2} + \frac{(S_p - S_a)^2}{\frac{1}{n}\Sigma(P_i - A_i)^2} + \frac{2(1 - r) S_p S_a}{\frac{1}{n}\Sigma(P_i - A_i)^2} \qquad (4)$$

Figure 2. The Sum of the Squared Errors

Analysis of the Data

Table 1 presents data about the frequency of the forecasting errors. Table 2 presents the coefficients and is followed by a discussion of their meaning.

As Table 1 shows, the real growth rate used in DoD's resource projections exceeded the real growth rate realized in the amounts appropriated in 66 of 94 fiscal years (70 percent of the projections). The effects of inflation have been removed from both the resource projections and the appropriated amounts. The optimistic tendency (70 percent of the projections exceeded the actual appropriations, in real dollars, vice the approximately 50 percent in an unbiased system), therefore, is not a result of the difficulty of forecasting inflation rates.

Bias

The bias proportion represents deviations in central tendency. It shows the proportion of the root mean square error that results from the difference between the mean of the predictions and the mean of the realizations. Positive values for the difference in the means of the predicted and realized values indicate that, on the average, higher real growth rates are projected than are realized.

In five of the six administrations the mean prediction exceeded the mean realization. The importance of that bias is indicated by the bias proportions in Table 2. In each of President Reagan's administrations, about 75 percent of the error in projections derived from optimism about how much Congress would appropriate. In President Bush's

Table 1. Frequency of Forecasting Errors

Number of periods forecast	94
Forecast real growth rate exceeded actual rate	66
Actual rate exceeded forecast	28

Table 2. Inequality Proportions

Administration	Bias	Variance	Covariance
Ford	.00322	.23899	.75779
Carter	.33380	.32189	.34431
Reagan I	.75237	.03787	.20976
Reagan II	.75249	.00044	.24707
Bush	.48722	.05170	.46109
Clinton	.26872	43326	.29802

administration, about 49 percent of the projection error resulted from an upward bias. About 27 percent of the projection error in President Clinton's first two years resulted from overly optimistic projections. President Ford's administration exhibited very little bias. About .3 percent of his projection error resulted from general overoptimism.

In contrast, President Carter's administration exhibited a bias below what the Congress appropriated, accounting for about one-third of the projection error.

VARIANCE

The variance proportion is zero only if the standard deviations of the projected and realized real growth rates are the same. As Table 3 indicates, for the administrations of Presidents Ford, Carter, and Bush and for President Reagan's first administration, the variance of the realizations exceeded the variance of the projections. For each of those administrations, the projected real growth rate fluctuated less from year to year than did the achieved real growth resulting from Congressional appropriations. For Presidents Reagan and Bush, this difference in consistency contributed only about 3.8 percent and 5.2 percent, respectively, of their projection error, making that source relatively unimportant compared to the effect of the upward bias in central tendency. During the Carter and Ford administrations, the difference in consistency was relatively more important, contributing 32 percent and 24 percent, respectively, of the projection error.

Conversely, in President Reagan's second administration and in the first two years of President Clinton's administration the projections have been less tightly distributed than have the congressional appropriations.

One might hypothesize that the pattern of the variance relationship indicates that Presidents Ford, Carter, Bush, and Reagan (in his first term) had a better-defined vision, or at least a firmer vision, for the national security than did the Congress. Such a hypothesis would accept year-to-year consistency in appropriations as a proxy for a consistent vision. A full

Table 3. Means, Standard Deviations, and Correlations

Administration	$P - A$	$S_p - S_a$	r
Ford	.0028	–.0238	–.0736
Carter	–.0294	–.0289	.0820
Reagan I	.0522	–.0117	.7325
Reagan II	.4934	.0019	.0937
Bush	.0436	–.0141	.1458
Clinton	.0123	.0156	.8524

examination of that hypothesis, however, is beyond the scope of this paper; for the present, we leave it for others to address. As one reviewer noted, however, it might be addressed through an analysis of the concurrent resolutions on the budget.[5]

COVARIANCE

The covariance proportion is zero only if the coefficient of correlation is 1. As indicated in Table 3, the directional agreement, the correlation, between the administration's real growth projections and the congressional appropriations have been highest in President Reagan's first term and in President Clinton's first two years. President Ford and the Congress moved in opposite directions. The correlations in President Carter's administration, President Reagan's second administration, and President Bush's administration each are positive, but quite low. The difference in correlation contributed relatively significantly to the projection error in the Ford administration, less so in the Bush administration, and progressively less so in the Carter, Clinton, and Reagan administrations.

The preceding analysis addressed the sources of the projection error on a relative basis. If the projection error is small, then the importance of a relatively large proportional contribution also is small. Thus, it is important to address the size of the projection errors. Did the administrations have similar projection errors, or did some administrations experience notably large projection errors? What was the source of any larger-than-typical errors?

SIZES OF THE PROJECTION ERRORS

Table 4 presents the average sizes of the projection errors as a percentage of the planned resource level; that is, as a percentage of the projection. Importantly, for Table 4, the calculation is based on the planned resource level (in constant dollars), not on the year-to-year real growth rates, and is not represented

Table 4. Size of Resource Shortfalls[a]

Administration	Mean Shortfall as Percent of Planned Resource Level
Ford	–5.3
Carter	+10.9
Reagan I	–13.4
Reagan II	–19.5
Bush	–11.1
Clinton	–2.5

[a] A minus sign indicates available resources were less than planned. The comparison is across the periods projected during each Presidential term.

logarithmically. The resource-level base portrays the effect on program management better than do the calculations based on year-to-year rates.

It is the error in projecting year-to-year real growth rates that causes the resource shortfalls and that error is an accurate portrayal of the overoptimism. The overly optimistic projection of future resources derives from the overly optimistic projections of real growth. Nonetheless, once resources are realized, it is the resource quantity that constrains management of operations and investment. Hence, the importance of those shortfalls is better measured as a function of the resource levels. Note that measurement using the resource levels makes each year's error dependent on the cumulative effect of the prior years' errors, as it in fact is.

Of the six administrations, only President Carter's projected less in resources than were realized. It is enlightening, however, to look at the timing and circumstance of those in-excess-of-projected realizations. During President Carter's tenure, the Congress appropriated an average of 4.9 percent more than President Carter requested. The Carter administration projections for his post-tenure years, fiscal years 1982 through 1986, were significantly less than Congress appropriated for those years. President Carter's plans for those post-tenure years were overfunded by an average of 13.8 percent of those plans.

Thus, President Carter's average resource overrun of 10.9 percent compared to his out-year projections can be attributed largely to President Reagan's military buildup. The Carter administration's bias to the low side of those realizations appears to be a result of a changed national security policy and perception.

Conversely, the existence of overfunded plans during President Carter's tenure confirms that it is possible for a President to overcome the systemic fiscal overoptimism of the defense establishment. Of the six administrations analyzed, only the Carter administration presented requests to the Congress that were less than the

Table 5. Correlation of Error Size and Bias

Administration	Absolute Value of Mean Shortfall as Percent of Planned Resource Level (%)	Bias Coefficient (%)
Reagan II	19.5	75.25
Reagan I	13.4	75.24
Bush	11.1	48.72
Carter	10.9	33.38
Ford	5.3	3.22
Clinton	2.5	26.87

amount ultimately appropriated by the Congress. If we conclude that the defense budget process includes a systemic over-optimism, then we are led to conclude that President Carter managed to overcome that systemic bias.

Spinney (1996) offers a description of the pressures to increase budget allocations that the defense establishment can place on a President. His recounting of the pressures and maneuvering leading to the 1996 increases in the future-years program provides considerable insight into the difficulties a President faces in overcoming defense's tendency to optimistic out-year fiscal projections.

Correlation of Projection-Error Size and the Bias Coefficient

Those administrations having the largest projection errors, as measured by Table 4, also exhibit the largest bias coefficients. Consider Table 5. The apparently high correlation is confirmed by a Spearman Rank correlation test. That test, yielding a rank correlation coefficient of .94, is significant on a one-sided test with a type I error of .02.

Thus, over the past 20 years, those administrations that exhibited significant bias (optimistic or pessimistic) in their resource projections, tended also to have relatively large errors in their projections of resources.

Consider Table 6, which is Table 2 reordered from the largest to the smallest projection error, except for the Ford administration. If we accept that President Ford's projection error derived primarily from his directional differences with the Congress, then the evidence becomes more persuasive.

If the bias coefficient is large, then the average predicted change is substantially different from the average realized change. If bias remains a major source of error over time, then the forecasting system is not improving. That is a serious error. The covariance error source should not be expected to approach zero. Were that true,

Table 6. Inequality Proportions Ordered by Size of Projection Error (Except for Ford Administration)

Administration	Bias	Variance	Covariance
Reagan II	.75249	.00044	.24707
Reagan I	.75237	.03787	.20976
Bush	.48722	.05170	.46109
Carter	.33380	.32189	.34431
Clinton	.26872	.43326	.29802
Ford	*.00322*	*.23899*	*.75779*

the line of predictions and realizations would be straight. Such an exact alignment is too much to expect (Thiel, 1971, p. 32).

Efforts to Correct the Bias Error

Bias has been recognized as a serious source of error by an independent commission and within the Department of Defense (Packard, 1986; Gansler, 1989; Spinney, 1996). Recognition within the Department is difficult to document because internal DoD management and financial management policy analyses are not publicly available. Nonetheless, there have been sufficient occasional recognition of fiscal overoptimism as a management problem to support the conclusion that the professional career staff was aware of it and of its deleterious effects (Lewis, 1994; Clark, 1990a; Clark, 1990b; Jordan, 1990).

"Bias has been recognized as a serious source of error by an independent commission and within the Department of Defense"

The Packard Commission (1986) focused intensively on the tendency to overestimate the future resources as a serious management problem. That Commission's report, together with pressure from career executives, fostered a limited recognition within DoD of the need to improve the forecasting of resources.

Conclusions and Recommendations

A tendency exists for the Defense Department to project the availability of significantly more resources than become available. Historically, those administrations having demonstrated the greatest bias in their real growth projections also have experienced the greatest shortfalls in resources. Hence, those administrations having demonstrated the greatest bias in their real growth projections also most seriously handicapped program managers. Projecting significantly more resources than become available directly affects force mix and capability. The force-mix optimization studies used in programming decisions incorporate a resource constraint.

The existence of optimistic bias has spanned administrations. It continues despite changes of administrations—whether the political party of the incoming administration is the same or changes. It appears, therefore, that the bias results from some characteristic of the defense management system; it is a systemic phenomenon. So it appears reasonable to conclude that reducing the optimistic bias will require changes to the planning and budgeting system. In undertaking such changes, it is important to recognize that bias reduction is the goal, not elimination of the projection error.

There clearly is room for improvement in the Defense planning and budgeting system. The analysis in this paper is empirical. It establishes existance of a systemic bias in one of the nation's major accounting and budgeting systems. Gansler (1989) and Clark (1990a; 1990b) each have identified significant costs

arising from budget turbulence in DoD. The systemic bias identified here is a source of that turbulence. It seems reasonable to suggest that other analysts could contribute significantly to the national welfare via rigorous development of improved forecasting methods that would be unbiased. A broad proposal for such research is outlined below.

Changes in the planning and budgeting system to reduce optimistic bias should be based on a review that identifies the decision points and techniques of the system. Techniques include the modeling and projection methodology; for example, regression analysis, auto regressive integrated moving average (ARIMA), or dynamic economic models. Decision points are those places in the process where out-year assumptions are made. Examples of these are whether DoD will receive a greater or smaller share of the U.S. budget, whether the U.S. budget will increase or decrease, and the size of the applicable growth rates.

Because the analysis identifies a period in which the systemic bias was corrected, a comparison of that period to other periods appears potentially fruitful. The first step in such research might be structured interviews with senior officials and analysts who played key roles in the planning and budgeting process under the Carter administration and other administrations.

Three sources of projection error were identified: bias, variance, and covariance. It is reasonable to expect that forecasting systems should exhibit the ability over time to diminish the bias source. Not to do so indicates lack of continuing improvement in the forecasting system. The time trend of bias errors does not indicate any systemic improvement. From a system perspective, the national defense planning system is not functioning as it should.

The variance error source appears to result from the relative consistency of the administration's vision of the national defense versus the consistency of the Congress's vision. Testing and analysis of that hypothesis is deferred, but changes to the forecasting system appear an unlikely way to improve the correlation of the Administration's and the Congress's vision for national defense.

The covariance error source should be expected to continue; further, improvements in the forecasting system that reduce the bias source almost surely will increase the relative size of the covariance error source.

Dr. Leland G. Jordan is an associate professor in the School of Business at Christopher Newport University, Newport News, VA. A retired member of the Senior Executive Service, Dr. Jordan has a DBA from George Washington University, a MS (Systems Analysis) from the Air Force Institute of Technology, and B.S. from the University of Florida. He has published in a number of academic and professional journals.

(E-mail address: ljordan@cnu.edu.)

REFERENCES

Chapman, J. (1981, summer). Fiscal stress and budgetary activity. *Public Budgeting and Finance,* 83–87.

Clark, R. (1990a). *The dynamics of budget instability* (research paper). Ft. Belvoir, VA: Defense Systems Management College.

Clark, R. (1990b). *Defense budget: A costly malpractice* (research paper). Ft. Belvoir, VA: Defense Systems Management College.

Department of Defense. (Various years). DoD press releases on the DoD budget. Washington, DC: DoD.

Department of Defense. (Various years). Annual Reports from the Secretary of Defense to the Congress. Washington, DC: DoD.

Department of Defense Comptroller. (Various years). National defense budget estimates. Washington, DC: DoD.

Financial Accounting Standards Board (FASB). *1985 Qualitative characteristics of accounting information. Statement of financial accounting concepts No. 2.* Stamford, CT: FASB.

Gansler, J. S. (1989). *Affording defense.* Cambridge, MA: MIT Press.

Isaacson, W. (1983, March 7). The winds of reform. *Time,* 12–30.

Jordan, L. G. (1990). *Systemic fiscal over-optimism in defense management.* Presented to the Washington Accounting Research Society, Loyola College, MD.

Joyce, P. (1996). Congressional budget reform: The unanticipated implications for federal policy making. *Public Administration Review, 56*(July-August), 317–325.

Lewis, K. N. (1994). The discipline gap and other reasons for humility and realism in defense planning. In P. K. Davis (Ed.), *New challenges for defense planning.* Santa Monica, CA: RAND.

Packard, D. (1986, June). *Report to the President by the President's Blue Ribbon Commission on Defense Management: National security planning and budgeting.* Washington, DC: The Packard Commission.

Quade, E. S., & Boucher, W. I. (1968). *Systems analysis and policy planning: Applications in defense.* Santa Monica, CA: RAND.

Sessel, D. (1995, November 3). The art of tweaking: White House's altered forecast on economy underlies Capitol Hill budget duel. *Wall Street Journal,* p. A16.

Shick, A. (1996). The majority rules. *Brookings Review, 14*(Winter), 42–45.

Spinney, F. C. (1980, May). Defense facts of life. *Armed Forces Journal.*

Spinney, F. C. (1992, May-June). Uncle Sam's budget shambles. *Challenge!* 9–12.

Spinney, F. C. (1996, July-August). Defense time bomb: F–22/JSF case study hypothetical escape option. *Challenge!* 23–33.

Thiel, H. (1971). *Applied economic forecasting.* Amsterdam, The Netherlands: North-Holland.

ENDNOTES

1. The optimistic assumption is production quantity, not slope.

2. Optimization within resource constraints is well established in national security planning. The techniques and theory were set out 30 years ago by Quade and Boucher (1968). The Packard Commission's report (Packard, 1986) clearly reflects the continuation of that practice.

3. A Future Years Defense Program covers a specified 6-year period for which DoD plans. A new FYDP period starts each biennium, thus constituting a rolling coverage of the future.

4. Data are from DoD press releases, Secretary of Defense Annual reports to the Congress, and the National Defense Budget Estimates series published by the DoD Comptroller. Data for earlier years were not available. Although the "Historical FYDP" reaches back to fiscal year 1962, it does not present the original estimates. FYDP data are revised if appropriations change during the year and also to reflect actual obligations through time. In addition, documents presenting the original inflation forecasts are not available and such original projections are necessary to restate the out-year data in constant dollars.

5. Analyses of the congressional budget process are in Joyce (1996) and Shick (1996).

TEST AND EVALUATION MANAGEMENT REFORM: *ISSUES AND OPTIONS*

James D. Love

Can a change in the management structure of the Department of Defense's test and evaluation infrastructure make it more cost effective and efficient while retaining the responsiveness and the weapons quality of the present system?

The Department of Defense (DoD) acquisition process and its test and evaluation (T&E) subprocess produce the world's finest weapon systems, and it retains a reputation for responsiveness to military needs and acquisition of quality weapon systems and other items. Yet the constant pursuit of greater cost-effectiveness and efficiency leads to questions and analyses of whether a different management structure—such as a single DoD T&E organization—would better accomplish these needs and goals. To properly consider this question, here we will focus on the infrastructure and management that supports the T&E process rather than on the T&E process itself, which consists of the planning, provisioning, and conducting of tests together with the analysis and reporting of data resulting from those tests.

As a starting point to the discussion, it is worthwhile to look at the makeup of the T&E infrastructure. Changes and improvements cannot be appreciated or understood without knowing the basis from which the changes are originating.

The DoD T&E infrastructure consists of the Major Range and Test Facility Base (MRTFB), whose policy guidance document is DoD Instruction (DoDI) 3200.11, the latest version dated January 26, 1998.[1] "The MRTFB is part of the National Test Facilities Base and is a national asset that exists primarily to provide T&E information for DoD decision makers and to support T&E needs of DoD research programs and weapon system development programs" (DoD, 1998, para. 3.1.2 & 3.1.3 and enclosure 2). Within the MRTFB, there are 21 test activities whose management is performed by four

components and nine commands with oversight by the Office of the Secretary of Defense (OSD) (Sanders, 1999). Membership of the MRTFB is listed in Appendix 1.

The Defense Test and Training Steering Group (DTTSG), chartered by the Under Secretary of Defense for Acquisition and Technology (USD [A&T]), acts as a permanent organization to coordinate planning and actions with respect to the MRTFB. Membership consists of representatives from the T&E and training communities of OSD and the military components. "These T&E ranges, where several thousand test projects are performed each year for DoD, other federal agencies, U.S. allies, and commercial users, are worth $25 billion and account for more than 50 percent of the total DoD land area in the continental United States" (Cohen, 1998).

The T&E infrastructure accounts for about 1.6 percent of the total DoD infrastructure budget, about $1.85 billion in fiscal year 1997 dollars. The total acquisition infrastructure is approximately 9.1 percent of the DoD infrastructure (Institute for Defense Analysis [IDA], 1998). Funds flow to the T&E facilities through several accounts and Service components, the primary ones being research, development, test, and evaluation (RDT&E) accounts, operations and maintenance (O&M) accounts, military personnel, procurement, and military construction accounts for the Services. These funds are used to keep the facilities ready for customer use and for upgrades to capabilities.

The users of the facilities pay for the direct support provided to them, which amounted to $1.5 billion in fiscal year 1997 (IDA, 1998). The DoD infrastructure included 1,437,768 personnel in 1996. Of these, 18,845 people (1.3 percent) were assigned to T&E functions within the Services (IDA, 1998). Appendix 2 provides different perspectives on T&E funding.

The questions that are posed by the critics of the process focus on how efficient it is and whether there are ways in which the costs can be reduced while still maintaining quality. The critics believe the process is poorly managed, inefficient, too bureaucratic, and in need of reform (Sanders, 1999). Their criticisms can generally be summarized in the following six statements:

- The bureaucracy is too big and too complex.
- The bureaucracy suffers from excessive duplication.
- It does not provide for clear lines of command and accountability.
- It sustains a counterproductive incentive structure and limits the ability of acquisition executives to effect cultural change.
- It exacerbates the natural tendencies of the Services to favor parochial solutions.
- It has not responded to the post-Cold War decline in acquisition spending and manpower levels (IDA, 1995).

We must not forget that the U.S. defense acquisition process has produced the finest combat systems in the world in

spite of its inefficiencies. As Jacques Gansler stated in his book *Defense Conversion,* "America designs and builds the best weapons in the world. These weapons, however, cost too much (especially in the small quantities likely to be bought in the future), take too long to develop and produce, and are often unreliable and prohibitively expensive to operate and support" (Gansler, 1995).

There are several reasons why these issues have been and are currently under debate. The most obvious is the Congressional language contained in the National Defense Authorization Act for Fiscal Year 1998, (Section 912. Defense Acquisition Workforce), which states:

> (a) Reduction of Defense Acquisition Workforce. (1) The Secretary of Defense shall accomplish reductions in defense acquisition personnel positions during fiscal year 1998 so that the total number of such personnel as of October 1, 1998, is less than the total number of such personnel as of October 1, 1997, by at least the applicable number determined under paragraph (2).
>
> (2)(A) The applicable number for purposes of paragraph (1) is 25,000. However, the Secretary of Defense may specify a lower number, which may not be less than 10,000, as the applicable number for purposes of paragraph (1) if the Secretary determines, and certifies to Congress not later than June 1, 1998, that an applicable number greater than the number specified by the Secretary would be inconsistent with the cost-effective management of the defense acquisition system to obtain best value equipment and would adversely affect military readiness.

The Strom Thurmond National Defense Authorization Act for Fiscal Year 1999 (Sec. 907. Management Reform for Research, Development, Test, And Evaluation Activities) states:

> (a) Analysis and Plan for Reform of Management of RDT&E Activities. The Secretary of Defense, acting through the Under Secretary of Defense for Acquisition and Technology, shall analyze the structures and processes of the Department of Defense for management of its laboratories and test and evaluation centers. Taking into consideration the results of that analysis, the Secretary shall develop a plan for improving the management of those laboratories and centers. The plan shall include such reorganizations and reforms as the Secretary considers appropriate.

The complete sections (912 and 907) of these acts can be found in Appendices 3 and 4.

Since 1987, more than 150 studies have addressed the need for DoD to achieve operational efficiencies in its RDT&E infrastructure. Figure 1 shows the more significant studies that have taken place from 1988 through 1998 (IDA, 1998). Recommendations from these studies focused mainly on management inefficiencies and

less on infrastructure reductions. For example, the 1995 "Directions for Defense: Report of the Commission on Roles and Missions of the Armed Forces" identified many opportunities for DoD to integrate operational activities with duplicative missions in areas such as command, control, communications, computers, and intelligence rather than RDT&E infrastructure reduction (Government Accounting Office [GAO], 1998).

One of the more recent studies is the 1995 Base Realignment and Closure (BRAC) study. "The BRAC cross-Service group for test and evaluation analyzed the capacity of 23 activities that supported test and evaluation of air vehicles, electronic combat, and armament/weapons and identified about 495,000 test hours of excess capacity. However, the group did not set capacity reduction goals" (GAO, 1995). This 1995 BRAC Cross-Service Analysis is widely quoted for its statement that there is a 52 percent excess T&E capacity for air vehicle, electronic combat, and armament/weapons testing infrastructure (IDA, 1998).

The 1998 Department of Defense Report to Congress estimated a 23 percent overall excess RDT&E base capacity. This excess capacity was broken down by Service as: Army T&E and lab facilities, 39–62 percent excess capacity; Navy/Marine T&E and lab facilities, 18 percent excess capacity; and Air Force product center, labs and T&E excess capacity, 24–38 percent excess capacity (IDA, 1998).

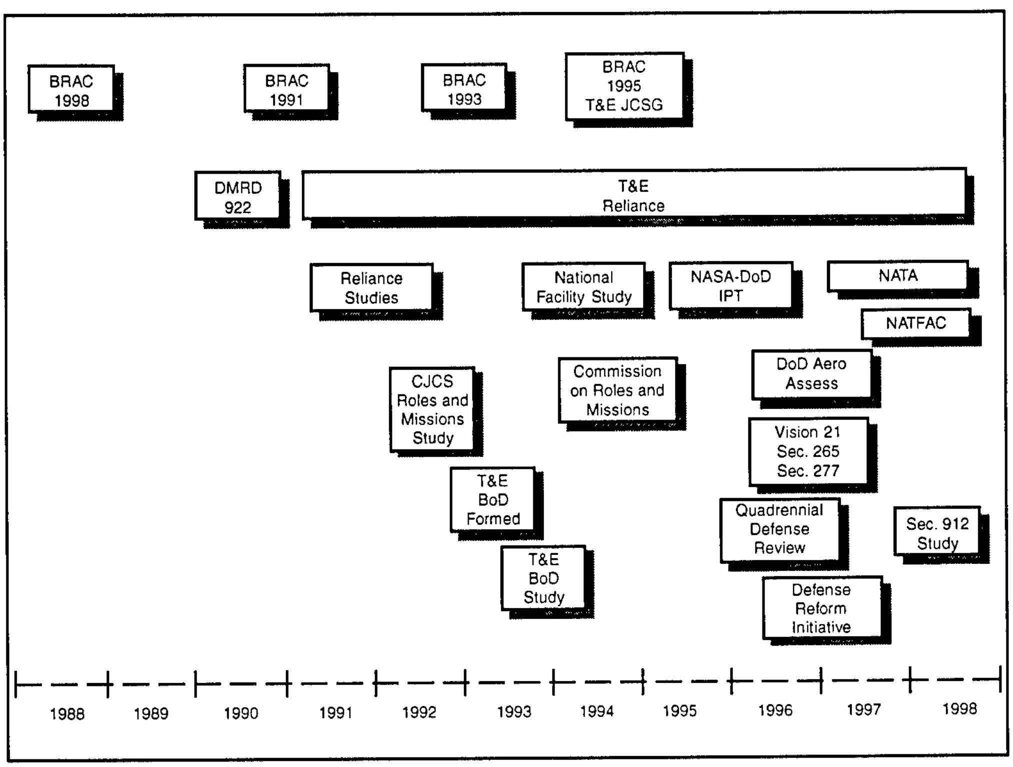


Figure 1. Studies Involving T&E

In addition to the need for reducing the amount of expenditures so that more is available for operational and modernization needs, there are also factors internal to the T&E community that drive a need for more efficient use of the available funds. The T&E infrastructure is aging, requiring costly repairs and upkeep; technology growth in the weapon systems being tested requires a commensurate growth in the measurement and evaluation capabilities of the T&E infrastructure, and a need for more flexibility and responsiveness to changing requirements all compel redesign of the infrastructure.

The current management structure evolved over time during the Cold War era and is complex and cumbersome with many oversight, coordinating, and approval links. While some recent changes in the OSD oversight structure have taken place,[1] the overall structure remains very complex, overlapping, and inflexible. It cannot respond either to short- or long-term market changes. Finally, since 1990, the funding has declined significantly, making it difficult to continue to support the MRTFB as it exists today. The user (customer) funded workload in work-years is down 25 percent from fiscal year 1990 to 1999 and institutional funding is down 30 percent from 1990 to 1999 (IDA, 1998).

DISCUSSION

The options for restructuring the T&E infrastructure management include simple changes that alter the current structure very little, to radical changes that would drastically change how business is done within the T&E community. Some of these options follow.

The first option would be to leave the current system in place as it is. This is a system of decentralized management. Each Service is responsible for its own facilities with an MRTFB framework of oversight. It is structured to provide high mission focus dedicated to the Services. A complex organizational structure, it consists of a number of committees, boards, councils, and steering groups (T&E Reliance Structure, Range Commander Council [RCC], DTTSG, Board of Directors/ Board of Operating Directors/ Joint Program Office [BoD/ BoOD/JPO], Test and Evaluation Resources Investment Board, Test and Evaluation Committee, and Service T&E Principals).

"The T&E infrastructure is aging, requiring costly repairs and upkeep...."

The MRTFB funding policy for direct cost to the users is somewhat self-regulating but lacks flexibility to accommodate special circumstances. While the current structure uses DoD-wide personnel, contracting, financial, management, and administrative practices, it still remains difficult to compare costs between facilities and ranges and especially between Services. The current system favors Service priorities over defense-wide priorities, making it difficult at times to get cross-Service support. Recapitalization of aging facilities is difficult to achieve and it is hard to make long-term commitments to potential users or to make long-term commitments to providers of services under the existing management structure (IDA, 1998).

A second option is that of combining each Service's T&E organizations into a single Service organization that would report to a Service T&E component commander. Included within the Service's T&E component organization would be not only the developmental testing organizations but also the operational test organizations and the Service's battle labs as well.

The T&E component organization could be placed under the Atlantic Command (ACOM). ACOM would have oversight over the Services' T&E activities but the policy and direct control would remain within the purview of the component commander. Funding for the operations, modernization, and support would derive from three sources. The users of the T&E facilities would continue to pay as they do now for the direct support provided to them. The Service's acquisition community, through a T&E funding program element, would provide the institutional funding.

"[A second option] allows the Services to retain control over the infrastructure and the funding for its operations and modernization, thus controlling its own destiny."

For example, the Air Force institutional funding line would be through the Assistant Secretary of the Air Force for Acquisition (AF/AQ). The Service operational side would provide the base operating allocation funds. This proposed structure would eliminate the need for policy level staff functions within the Service staffs (Engel, 1999). This option allows the Services to retain control over the infrastructure and the funding for its operations and modernization, thus controlling its own destiny. Oversight, however, shifts to the operational users rather than remaining in OSD as in the current structure.

A third option is that of a continental United States (CONUS) Range Command structure. The command would place the T&E activities within the MRTFB under the purview of a single commander who would report to the Board of Operating Directors, which is comprised of the Services' vice-chiefs. The O&M budgets would remain within the Service accounts. The focus would be on operational cost-efficiency. The ranges would be considered as parts of an integrated whole, rather than as separate facilities. Operations, under this proposed concept, would be contracted out under a single A-76 O&M contract.

The proposed plan would allow the ranges to be placed into and out of caretaker status with a 30-day call-up notification as the workload surges and shifts. The workforce would be shifted as required to meet the workload demands of the various ranges. This approach should result in significant cost-savings from work force reductions. There are a number of potential political issues that would have to be resolved for this concept to work, however. Since the Services retain possession and control of the ranges, it is believed that there are no legal barriers to implementing this option and it should not require congressional approval (Hollis, 1999).

The fourth option consists of realigning the funding to a centralized funding line for justification, appropriation, and distribution. User funding would not be

affected and would remain as it is today. Establishment of requirements and needed capacity for the next 10- to 15-year period would be done during each Quadrennial Defense Review through a joint Service, OSD, Defense Advanced Research Projects Agency, and Ballistic Missile Defense Organization working group. The requirements and capacity forecast would be provided to OSD, the Services, and BoD for developing the Future Years Defense Program. The DTTSG would have final approval on priorities and locations for the spending plans.

The execution would be through the Service channels with execution monitoring being done by the BoOD/JPO, who would provide semi-annual reports to DTTSG. The DTTSG/BoOD would be responsible for developing the program objectives memorandum and budget estimate submission inputs as well as responding to congressional issues. Annually, after the appropriation bill passed, OSD/DTTSG would approve distribution of funds to Services for execution. The Services would be responsible for distributing funds to the MRTFBs for O&M and investments. Charge policies for range use would be as specified in DoD regulations. The implementation of this option would require congressional approval (IDA, 1999).

The fifth option for consideration is establishment of a Defense T&E Command (DTEC) under the Defense Logistics Agency (DLA). The Services would retain the bases and support infrastructure while the T&E activities would be performed as tenant activities. Funding, both user and institutional O&M, would remain as it is now. A central account would provide funds for all investments and personnel support. The local Service commander would continue to operate the T&E activities. Military personnel would be assigned as they are now under the current system.

The DTTSG and BoOD would provide oversight to ensure that the Service and OSD priorities were recognized. DTEC would provide the day-to-day management and run the investment programs. DTEC would also be the test location recommendation source. There are several variants of this option which are: to include all MRTFBs, only the RDT&E MRTFBs, or facilities could be grouped by category of major focus testing such as aircraft, weapons, etc. This option would require congressional approval for implementation (IDA, 1999).

"A central account would provide funds for all investments and personnel support."

A sixth option is to consolidate all T&E infrastructure under the management control of a DoD T&E agency (DTEA). Funding and management would be by the agency in a single account for O&M and a single investments account. Day-to-day operational scheduling could be done either locally or by a single centrally located scheduling office. The latter would allow for more effective scheduling of joint multifacility programs.

The need for congressional approval of this option is a source of debate. Supporters contend that the Secretary of Defense has the authority to make this change in *U.S. Code* Title 10, which states, "Whenever the Secretary of Defense determines such action would be more effective, or

efficient, the Secretary may provide for the performance of a supply or Service activity that is common to more than one military department by a single agency of the Department of Defense."

A seventh option is to establish a Defense Acquisition Agency (DAA) that would combine all of the acquisition functions within each of the Services under a single DoD agency. T&E would become consolidated under a department within this agency. The functions and operational approach would be similar to the approach under the option above (IDA, 1999). "The concept of a centralized, civilian-operated weapons systems acquisition agency was considered during both the First and Second World Wars. However, all proposals for such an agency were rejected" (GAO, 1986).

The most serious recent discussion of this policy change was in 1986 when the GAO was required to look at a centralized, purely civilian acquisition organization by Congress in the National Defense Authorization Act of 1986. The report listed a number of advantages and disadvantages and stated, "The major acquisition problems most often described were: (1) inadequate requirements identification, (2) program instability, and (3) a lack of uniform policy implementation. The predominant views expressed were against the Agency. Many believed that any advantages offered would be more than offset by the disadvantages" (GAO, 1986).

A RAND study assessed the push for centralization this way (Donohue, Lorell, Smith, and Walker, 1993):

> During the last months of the Bush administration, high-level decision makers discussed the possibility of consolidating all military R&D and acquisition into a single civilian DoD agency, with additional DoD agencies for Science and Technology, and Test and Evaluation. Under such a plan, the military Services would still generate weapon system requirements, but from then on all R&D, development, and testing would be the responsibility of the centralized civilian agencies. The goal of centralization would be to reduce overhead, improve management, eliminate duplication, increase economies of scale, and tighten control to minimize cost growth and schedule slippage.

The GAO report, however, was not optimistic about the results (1986):

> For years, advocates of greater centralization of the U.S. acquisition process have pointed to the highly centralized civilian acquisition bureaucracies of many of our major allies in Europe and elsewhere as possible models. Probably the foreign model most often mentioned is the French system, which is dominated by the centralized acquisition agency called the Delegation Generale pour l'Armament, or DGA. Are these foreign centralized agencies indeed more efficient? Unfortunately, there is little reliable data to indicate clearly that foreign organizations manage their limited military R&D resources more efficiently.

The eighth and final option is a uniquely different approach. This option proposes to establish a government corporation that would be a semigovernmental entity chartered by the government to manage and provide T&E services. This approach is used extensively by state, local, and federal governments. The *U.S. Code* (Title 31, Subtitle VI, Chap. 91, § 9102) governs the application of government corporations at the federal level.

Some of the more familiar examples of using this approach in state and local governments are turnpike authorities, water and sewer commissions, and airport authorities. Examples of use at the federal level are the Tennessee Valley Authority, Federal Deposit Insurance Corporation, AMTRAK, and the Saint Lawrence Seaway Development Corporation.

Its past application has been for noncore functions, which government has a legitimate interest in seeing are performed in support of the general public. The use of this approach should allow the government to implement the best business practices of the private sector in personnel management, contracting procedures, long-term commitments to service providers and service recipients, and allow long-term investments using best commercial practices while preserving government interests. The amount of control exercised by the government can be set at any level desired by the government. Implementation of this option would require congressional approval and would likely face tough challenges in the political arena. The option could be exercised at several different levels, including strictly T&E facilities or, at a broader level, including other RDT&E facilities.

Established to operate like a commercial enterprise, it would operate using best commercial practices and be managed by a CEO from the private sector, who is compensated as a private-sector CEO. Oversight and control would be afforded to the government through a Government Board of Directors (GBOD) comprised of government and private sector members. Such a government corporation would have the authority to float bonds for investments, just as a privately owned corporation would (GAO, 1986). This option provides the flexibility needed to adjust to market and cultural changes.

CONCLUSION

The options discussed above represent only a sampling of those that are possible. They do, however, represent the thinking of some of the most senior and most experienced minds in the test and evaluation business and represent a cross-section of the thinking within the Services, OSD senior staff, and the defense industry. The options represent widely diverse positions that range from minor changes within the management structure to radically altering the structure.

"A central account would provide funds for all investments and personnel support."

To assess the options, a reference framework is needed. The Services prefer to remain in control of their own destiny, which means they desire to retain as much control as possible. From an overall DoD perspective, it is desirable to have the least

duplication, most cost-efficient and effective possible—a solution optimized for the joint overall good. Since the current structure represents the Service control option and, if the critics are correct, has not led to an effective and efficient structure, an alternative form of management structure would seem to be justified.

Table 1 summarizes the framework used to compare the options. Since option 1 is the current structure and has been discussed previously, only the other seven options are shown in the table. A discussion of charge policies is omitted from the table since it is the same for all except option 8. The last row of the table (Row J) captures the concerns of the critics about the current structure and the capability of each of the options for addressing those concerns. The note in Table 1 lists the criticism discussed above and ranks the criticisms in importance based on the current structure's weaknesses. That ranking is as follows:

"Option 2, Service T&E Commands, addresses several of the issues but not the three issues of most concern: market flexibility, adaptability to cultural change, and the responsibility for making the hard decisions."

- **Market response (short-term flexibility):** How fast can the T&E management structure respond to changes in the marketplace?
- **Cultural changes (long-term flexibility):** How rapidly can the T&E management structure respond to changes in the acquisition culture such as the end of the Cold War?
- **Clarity of the command chain/ responsibility:** How clear is the chain of command and identification of the responsible individuals who should be making the difficult decisions?
- **Parochial (Service-focused) solutions:** Strongly correlated to the command chain/responsibility aspect. Are the decisions made from a parochial view or are they made from a joint/ DoD perspective?
- **Duplication:** How well can the management structure assess and respond to duplication of facilities?
- **Bureaucracy:** How "sluggish" and overpopulated is the management structure? This was given the least priority in the ranking system because it is tied to several of the issues above. If the management structure is open to cultural changes and has a clear command chain that is responsible for acting, then the bureaucracy can be managed quite well.

Option 2, Service T&E Commands, addresses several of the issues but not the three issues of most concern: market flexibility, adaptability to cultural change, and the responsibility for making the hard decisions. This option does have an advantage over the current structure in eliminating the need for policy level staff functions within each Service staff, thus reducing somewhat the manpower in the oversight role, but this is not a significant enough advantage to warrant change.

Characteristics	Option 2, Service T&E Commands	Option 3, CONUS Range Command	Option 4, Centralized Funding	Option 5, Defense T&E Command (DTEC)	Option 6, DoD T&E Agency (DTEA)	Option 7, Defense Acq. Agency with T&E Department	Option 8, Government Corporation
A. Management	———	———	———	———	———	———	———
1. Strategic level	Service T&E commands	CONUS command	Service	DTEC	DTEA	DAA	Government Board of Directors (GBOD)
2. Day-to-day operations	Local commander	Local commander	Local commander	Local commander	Local supervisor	Local supervisor	Local supervisor
3. Congressional interface	Service	CONUS command	DTTSG/BoOD	DTEC, Services	DTEA	DAA	CEO
4. PPBS responsibility	Service	CONUS command	DTTSG/BoOD	DTEC, Services	DTEA	DAA	N/A
5. Management focus (mission, efficiency, etc.)	Mission	Mission	1. Mission 2. Investment efficiency	Efficiency	Efficiency	Efficiency	Efficiency and effectiveness
B. Funds management	———	———	———	———	———	———	———
1. O&M	Service	CONUS Command	DTTSG/BoOD	Services	DTEA	TED	Corporation
2. Investments	Service	CONUS Command	DTTSG/BoOD	DTEC	DTEA	DAA	Corporation with GBOD approval
C. User/customer funds	User/program offices	User/program offices	User/program offices	User/program offices	User/program offices	User/program offices	User/program offices
D. Oversight	ACOM	BoD	OSD	DTTSG/BoOD	OSD	OSD	GBOD
E. Facility/range control	Services	Services	Services	Services	Services, host bases; DTEA, T&E facilities	Services, host bases; TED, T&E facilities	Services, host bases; Corp., T&E facilities
F. Users/customers (Note: All options place the DoD as top priority for customer support)							
1. Commercial	Yes, on non-interference basis	Yes, on non-interference basis	Yes, on non-interference basis	Yes, on non-interference basis	Yes, on non-interference basis	Yes, on non-interference basis	At commercial rates
2. Other	As allowed by law	As allowed by law	As allowed by law	As allowed by law	As allowed by law	TBD	At commercial rates
G. Personnel	———	———	———	———	———	———	———
1. Civilian Responsibility	Civil Service	Civil Service	Civil Service	Civil Service	Civil Service	Civil Service	Industry practices
2. Military	Service assigns in career fields	By career field from services	By career field from services	By career field from services	Technical advisors only	Technical advisors only	N/A
H. Implementing Authority	U.S.C. Title 10	U.S.C. Title 10	Congressional approval	Congressional approval	Congressional approval	Congressional approval	Congressional approval
I. Implementation Impediments	Political sensitivities	Political acceptance	Political acceptance	Political acceptance, Service resistance, multiple budgets	Political sensitivities, Services' resistance	Political sensitivities, Services' resistance	Political sensitivities, Services' resistance
J. Issues addressed by option*:	Cost efficiencies, issues 4,5, & 6	Issues 3, 4, & 5	Issues addressed by option; 4, 5, & 6	Issues addressed: 3, 4, 5, and 6	Issues addressed: 3, 4, 5, and 6	Issues addressed: 3, 4, 5, and 6	Issues addressed: 1, 2, 3, 4, 5, and 6

*: 1. Market response (short-term flexibility); 2. Cultural changes (long-term flexibility); 3. Command chain/ responsibility; 4. Parochial solutions; 5. Duplication; and 6. Bureaucracy.

Table 1. Test and Evaluation Restructuring Options

Option 3, CONUS Range Command, closely resembles option 2, but is one step closer to centralized control and to reducing dramatically the infrastructure manning through an A-76 action. Still, it does not address either of the top two issues of short-term flexibility (market response) or long-term flexibility (cultural change).

Option 4, Centralized Funding, is a financially focused approach that is centered on controlling the flow of investment funds as a way of controlling the future development of facilities and ranges and thus reducing duplication. It is limited in the changes it would be able to accomplish and would be slow in developing a solution. It would, however, probably gain eventual acceptance within the Services. It faces strong political opposition and requires Congressional Approval, a large battle to be fought for such a small gain since it does not address the three issues of most concern—market flexibility, adaptability to cultural change, and a clearly identified decision maker responsible for making the tough calls.

"If DoD is serious about revolutionizing the approach taken to the business of weapons procurement and testing, then it requires serious changes to the current management structure."

Option 5, Defense T&E Command (DTEC), is a variation of option 3. It suffers from a split in the ownership of the budget accounts between the Services and the command, a difficult situation to overcome. Although these accounts are separate, for purposes of justification and support, they are more easily supported if they come from the same organization (Service, command, etc.).

Option 6, DoD T&E Agency (DTEA), and option 7, Defense Acquisition Agency with T&E Department, are very similar. Both address a number of the efficiency and cost-effectiveness issues but are weak in providing the flexibility and market response agility that is desirable for rapid market responses and for overall long-term flexibility.

Option 8, the government corporation, is the only option that clearly addresses the two major issues of short-term and long-term flexibility. Many of the best practices that are desired from the acquisition community and its T&E subcomponent require changes to the way business is done—the "revolution in business affairs."

If DoD is serious about revolutionizing the approach taken to the business of weapons procurement and testing, then it requires serious changes to the current management structure. Only option 8 proposes those changes that will make the management structure responsive to the market because it is the only option that allows implementation of market response mechanisms.

This option also provides other desirable management structure characteristics such as a broad strategic span of control, minimized stovepipes, and delegation of authority to the lowest level possible that add to its appeal as a new way of doing business. Such a structure would minimize bureaucracy and flatten the organizational structure. Flexibility in personnel, contracting, and financial management practices would also be achieved. The nongovernment corporation would also allow

for more innovative financial business practices that would spur investments as well as provide for easier access to commercial services and more easily accommodate purchases from commercial or academic sources.

It, however, calls for changes that will be difficult to implement because of the drastic changes in the culture and control of resources necessary to make the change. The Services will resist, congressional approval will be difficult to achieve, and the current civilian organizational components will resist implementation. But, if Congress is truly interested in the DoD operating like a business, then it must support the organizational changes necessary to place those components (like the T&E infrastructure) that are most like the commercial world in a business-styled structure. In spite of these obstacles, option 8's potential benefits are worthy goals that should be pursued.

James D. Love is a 1993 graduate of DSMC (Class 93-2) and holds a Master of Science Degree in National Resource Strategy from the National Defense University (Industrial College of the Armed Forces, 1999.) He is also a 1999 graduate of the Defense Acquisition University's Senior Acquisition Course and was the recipient of the 1999 DAU Research Award. Mr. Love has extensive experience in test and evaluation serving since 1975 in both the field and in the Office of the Under Secretary of Defense for Acquisition and Technology, Test and Evaluation. Mr. Love also holds a Master degree in Business Administration from the University of West Florida and a Bachelor of Science degree in Electrical Engineering from the University of Alabama. Mr. Love is Chief of the Strategic Plans Division, Air Armament Center at Eglin AFB.

(E-mail address: lovej@eglin.af.mil)

REFERENCES

Cohen, W. S. (1998). *1998 Annual Report to the President and the Congress.* Washington, DC: U.S. Government Printing Office.

Commission on Roles and Missions of the Armed Forces (1995). *Directions for defense: Report of the Commission on Roles and Missions of the Armed Forces.* Washington, DC: Author.

Department of Defense (1998, January 26). Major Range and Test Facility Base (MRTFB). DoDI 3200.11.

Donohue, G.; Lorell, M.; Smith, G.; & Walker, W. (1993, April). *DoD Centralization: An Old Solution for a New Era?* RAND Issue Paper.

Engel, R. (1999, January 14). Commandant, Industrial College of the Armed Forces, interview conducted at Fort McNair, Washington, DC.

Gansler, J. S. (1995). *Defense conversion.* Cambridge, MA: MIT Press.

Government Accounting Office (1986, November). *Advantages and disadvantages of a centralized civilian acquisition agency* (GAO Report No. GAO/NSIAD-87-36). Washington, DC: Author.

GAO (1995, April). *Analysis of DoD's 1995 process and recommendations for closure and realignment* (GAO Report No. GAO/NSIAD-95-133). Washington, DC: Author.

GAO (1998, January 8). *Best practices: Elements critical to successfully reducing unneeded RDT&E infrastructure* (GAO Report No. GAO/NSIAD/RCED-98-23). Washington, DC: Author.

Hollis, W. (1999, January 25). U.S. Army HQ, Director of Operational Analysis, interview conducted at Pentagon, Washington DC.

Institute for Defense Analysis (1995, February). *Notes on consolidating acquisition organizations* (provided by Charles Ackerman).

Institute for Defense Analysis (1999, January 18-19). *Alternative T&E infrastructure management models* (briefing to DSB Task Force on Integrated Test and Evaluation).

Institute for Defense Analysis (1998, November 17). *Analysis of T&E infrastructure* (IDA briefing to DSB Task Force on Integrated Test and Evaluation).

National Defense Authorization Act for Fiscal Year 1998. Pub. L. 105-85, § 912 (1997).

Sanders, P. A. (1999, February). Range Commanders Council briefing. Office of Director, Test, Systems Engineering, and Evaluation (Acquisition & Technology): Washington DC.

Strom Thurmond National Defense Act for Fiscal Year 1999 (1998). Pub. L. No. 105-261, § 907.

U.S.C. Title 10, Subtitle A, Part 1, Chap. 8, Subchap. I, § 191(a).

ENDNOTES

1. On June 7, 1999, the Secretary of Defense approved the transfer of key test and evaluation responsibilities from the Office of the Undersecretary of Defense for Acquisition and Technology (OUSD[A&T]) to the Director, Operational Test and Evaluation (DOT&E). The SECDEF also directed that DODI 3200.11 be revised appropriately to reflect the realignment of responsibilities for the MRTFB and to reflect that DOT&E will establish policy for and composition of the MRTFB.

MANAGEMENT: *TOWARDS A UNIFIED FIELD THEORY*

Neal Pollock

Leadership and the One Minute Manager. By Kenneth Blanchard, Patricia Zigarmi, and Drea Zigarmi. William Morrow & Co., New York, 1985.

Executive Leadership. Elliot Jaques. Cason Hall, Arlington, VA, 1991.

Requisite Organization. Elliot Jaques. Cason Hall, Arlington, VA, 1989, (2nd ed. 1996).

This article is a "theme" review of three books written by management theorists of high repute. (It follows in the footsteps of a piece published here 4 years ago [Frisch, 1995].) These authors approach management from very different perspectives. But to quote Frisch, "Unfortunately, every con-ceptualization and interpretation needed to arrive at a theory represents the point of view or, more generally, the value system of the observer, and objectivity is just an illusion. The same 'facts' can have different meanings for different people; and even the same people may view the facts differently depending on the time and situation" (Frisch, 1995).

Fads in management seem as common as those in fashion. Their short life cycles may, however, be attributed not only to deficiencies in logic or implementation, but to the lack of a unified theory of management. Systems engineering teaches us that optimizing the whole de-optimizes the parts, and optimizing the parts de-optimizes the whole. Similarly, reengineering seems to indicate that some processes need to be replaced by revolution rather than updated by evolution. Perhaps a piecemeal approach prevents many interventions from reaching the required critical mass. Focusing on leadership, management, or supervision, and disregarding the other levels of abstraction or resolution, may be at the core of the difficulty.

Both authors attempt to expand initial, breakthrough concepts into management systems and to address, it seems, the

essence of management. The two systems are qualitatively diverse to say the least. Nevertheless, they are not antithetical.

SITUATIONAL LEADERSHIP

Blanchard developed "situational leadership" (and version II) some time ago. It was offered by the Navy in Crystal City until the Base Realignment and Closure Act moved much of the personnel out of Northern Virginia. In fact, they also offered his "leadership bridge" (Good, Hill, and Blanchard, 1992) which combined "situational leadership II" with the Myers-Briggs Type Indicator (MBTI) (Briggs-Myers and McCaulley, 1985). Of course, his most well-known work centered upon *The One Minute Manager* (Blanchard, 1991). In the present volume, he and his co-authors have combined situational leadership "II" with one-minute management. Perhaps his next endeavor will combine all three into one unified approach to management. It seems to me that the two components of the present volume dovetail quite well. I will not elucidate the principles of one-minute management here since, I believe, they have entered the mainstream of management practice; they center on concise and timely feedback (both positive and negative) between a manager and a subordinate.

"Some have said that managers manage things not people, implying that superiors are supervisors, not managers."

Situational leadership II, however, is a much more complex and far-reaching theory. The heart of it lies in the obvious observation that people differ. The extension to the Myers-Briggs is, thus, a very logical one. The latter is a model of 16 types of preferences to which each person belongs.

But Blanchard is not plumping for the nature versus nurture polarity. Rather, he elucidates that competence is a function of nature, nurture, and attitudinal factors, as applied to a task. Thus, for instance, a concert pianist requires not only natural talent and quality training, but also desire, drive, and confidence in order to succeed. This is a refreshing departure from the linear (let alone binary) assumptions of others, in practice if not in theory. Thus, Blanchard endows his model with four developmental levels (D1 through D4). His is a nonlinear formulation despite the appearance of linearity in these levels (see Figure 1). It then follows that persons at each developmental level require different treatment by their managers.

I use the term "managers" for simplicity. Many organizations (but not Jaques) would call them supervisors. Some have said that managers manage things not people, implying that superiors are supervisors, not managers. This is nonsense. An organization can *only* manage people. Machines (at present) cannot think, and thus cannot be managed. Furthermore, people primarily operate mentally. Organizations operate through their people. They can only operate with or upon things.

For our purposes, people are not things. Any student of social psychology should know that the subjects of their experiments cannot be apprised of the nature of the experiment without destroying its validity. People do not behave as automatons.

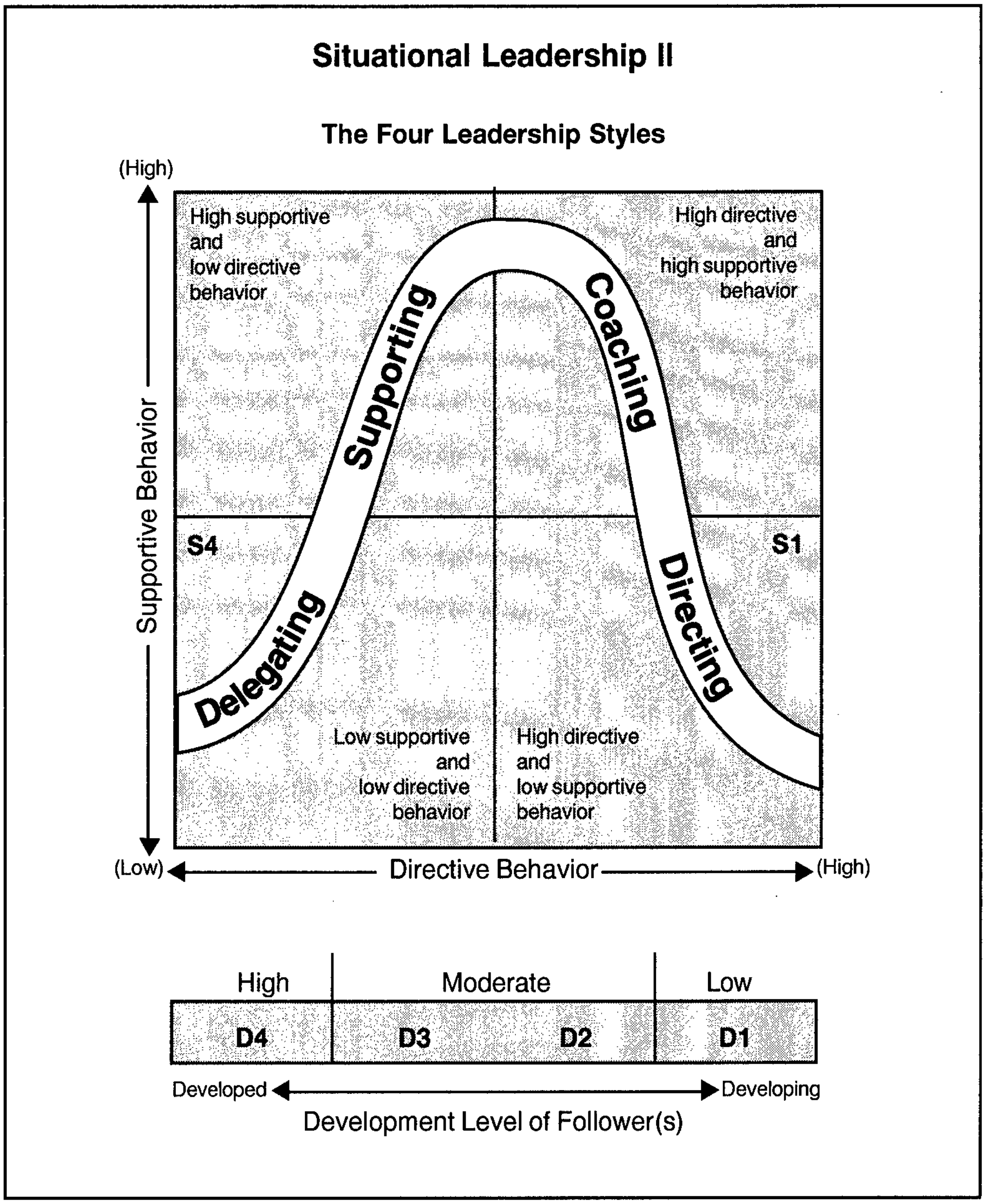


Figure 1. Blanchard's Management Model

Indeed, the researches of Freud and Jung have adequately demonstrated the existence of human nonrationality and, sometimes, irrationality—for those who have not noticed this themselves. There are precious few (if any) Commander Spocks or Commander Datas in this world. Notably, neither one was human; it is doubtful they would have been as endearing if they had been depicted as such.

SUPPORT AND DIRECTION

The two types of treatment devised by Blanchard are *support* and *direction.* These form a two-by-two matrix with each being high and low. The odd thing is that D1 requires only direction, D3 requires only support, D2 requires both, and D4 requires neither. It seems logical that direction would fall as a worker becomes more experienced. It is not, however, intuitive that support needs would go up, then down. But if one were to map "direction" onto the physical world as a linear process (which *seems* right) but map "support" onto the psychological world (wherein all human traits are normally distributed via the bell-shaped curve), the result would match Blanchard's model to a "T." Indeed, the "normal" or Gaussian shape (see Figure 2) (Cascio, 1978, p. 81) of Blanchard's curve is apparent in Figure 1.

Of course, Blanchard's resulting four styles of management are labeled S1 through S4, to match the developmental levels (S1 being generally appropriate to D1, etc.). He does provide for emergency contingencies, though. This is reminiscent of other models that calculate one's management style, but also predicate a backup style for emergencies.

Interestingly, the authors do not address any personal propensities in this regard for those practicing situational leadership. As the Myers-Briggs practitioner would do, one is, apparently, to develop other operating modes than one's natural, preferred one. Thus, it is assumed that the manager is proficient in the use of all four

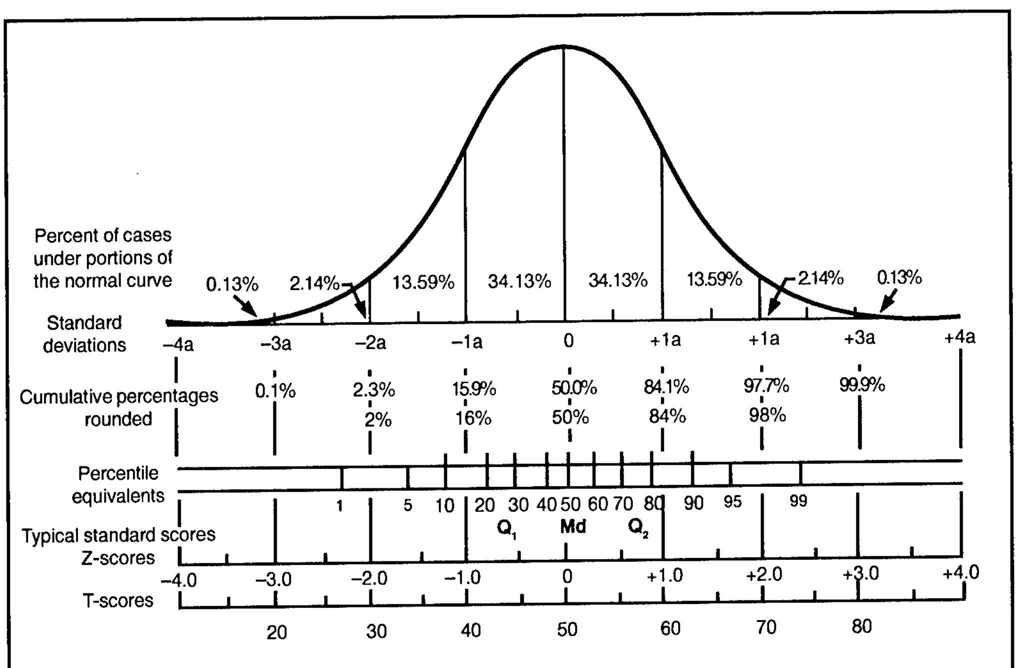


Figure 2. Normal or Gaussian Distribution

management styles. As is pithily stated, "there is nothing so unequal as the equal treatment of unequals" (Blanchard, 1985). Actually, Blanchard has merely elevated equality to a higher level of abstraction. Most paradoxes and enigmas are really confounded levels of abstraction. Similar to the systems engineering dictum cited above, what is true at one level of abstraction (or level of detail or level of resolution) is not true at all higher levels.

But Blanchard's approach is eminently practical. The small volume itself is extremely readable. Its stepwise approach is gently convincing and disarming. His style demonstrates his stylistic recommendations; it's a combination of coaching and counseling approaches to teaching. Indeed, it supports his multifaceted approach to people management: physical and psychological. In his own words, "Situational leadership is not something you do to people but something you do with people" (Blanchard, 1985). Such a low-key approach may make his techniques more palatable to a jaded workforce.

Nevertheless, as Pritchett points out in *High Velocity Culture Change* (Pritchett, 1993), organizational change requires a great deal of effort and commitment before it becomes institutionalized. When RADM John Ailes became Commander, Space and Naval Warfare Systems Command, he determined that his subordinate supervisory personnel required training in supervision. Despite the long-standing "requirement" that supervisors take five specific supervisory courses within a year of their appointment, precious few had completed the program and earned their "Supervisory Excellence Program" plaque.

RADM Ailes directed personnel to devise a short course and required *all* designated supervisors to take it, or he would remove them as supervisors. (Needless to say, we all took the course.) Since it was directed from above, the trainer asked, at the start of the course, what people thought about their being directed to take it. I still remember the response of one supervisor who said angrily, "I've been supervising for 30 years; I don't need this!" A voice in my head said, "yes, you've been supervising poorly for 30 years and you wouldn't want to change now!"

"Most paradoxes and enigmas are really confounded levels of abstraction."

Blanchard's method of operation includes contracting for leadership style. For this to work, both supervisor and subordinate must understand situational leadership. In addition, they should agree on the developmental level and matching style appropriate to the task at hand. Competence is, in this schema, transitive. There are no competent people, only people competent at a particular task. The task is the object. The developmental level can only be defined in terms of a given task. Since the task of supervision must also comply with this proviso, supervisors within the organization would be mapped onto appropriate levels of development and (their superiors') style.

Tasking Time-Span

Jaques's conception of development level is entirely different (see Figures 3 and 4). He claims that his empirical data

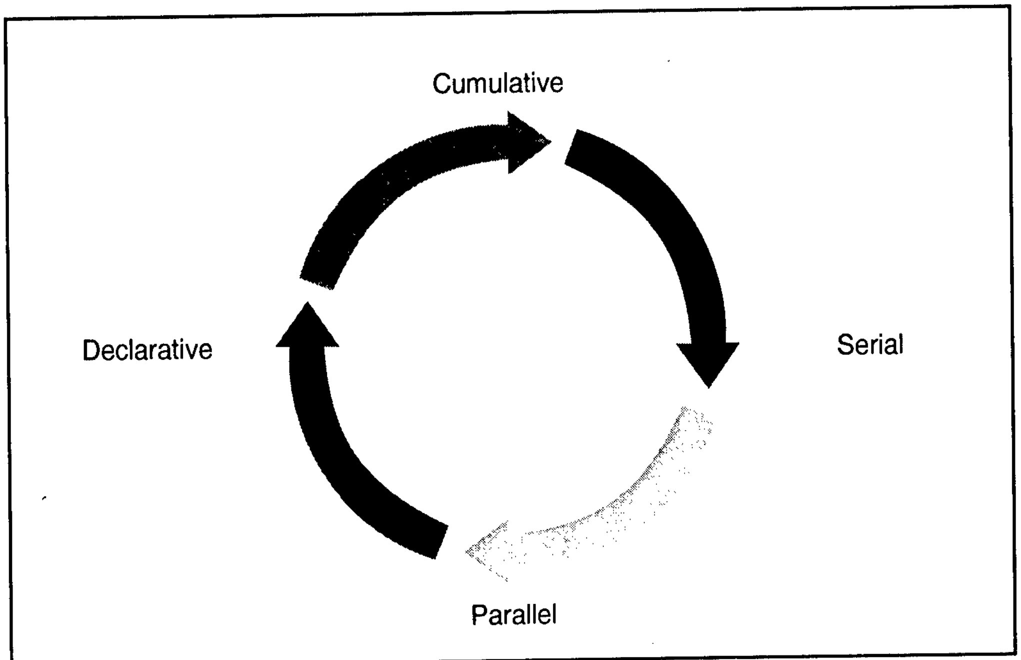


Figure 3. Jaques's Methods of Mental Processing

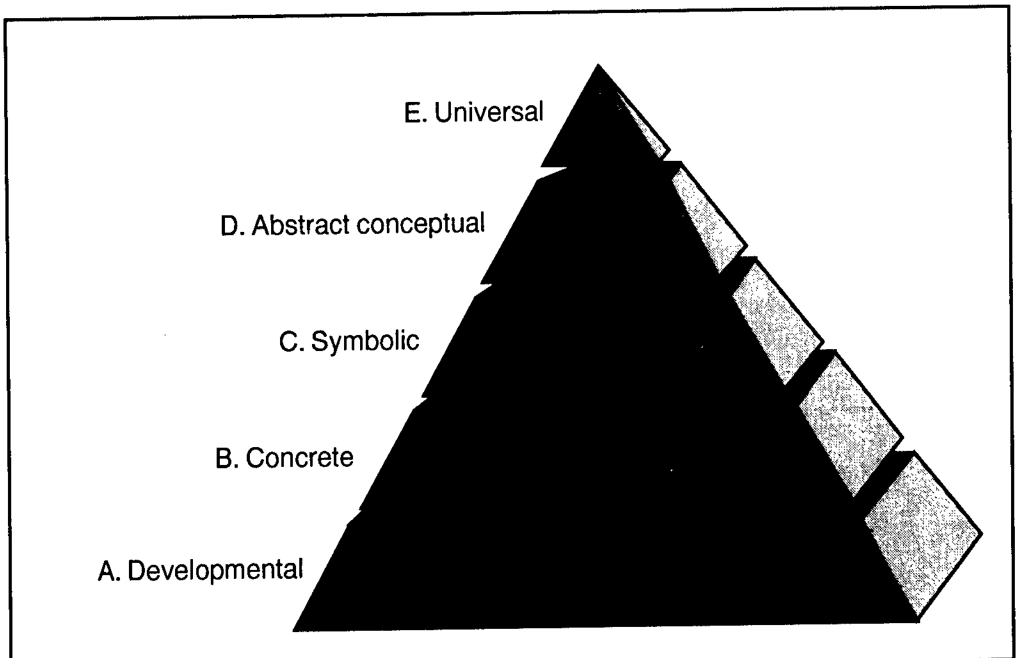


Figure 4. Jaques's Orders of Information Complexity

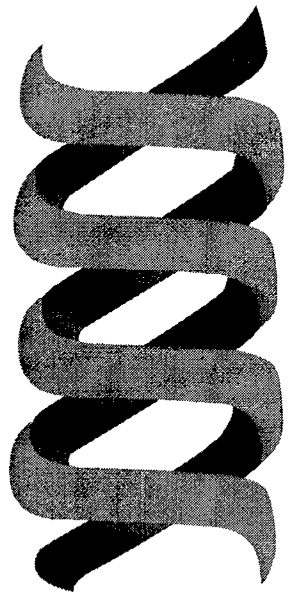


Figure 5. Cognitive Power

indicate that a person's maximum time-span of tasking determines his or her proper place in an organization. Indeed, he even correlates organizational rankings and (for a few selected countries) pay scales for these rankings. This time span or time horizon equals the maximum duration (from the very start to the very end) of one's longest task in a particular job. This is not a continuous scale, but has consistent break points (what we engineers call "knees on the curve"). Thus, he defines the optimal levels of an organization, not too vertical, not too horizontal, but naturally expressing human predilections for time slices. In another volume (*Time Span Handbook,* 1964) he provides specifics on how to determine the time-span of specific jobs.

To Jaques, competence involves matching an individual's time-span with a particular job's time-span. As does Blanchard, he finds competence to be transitive (i.e., a function of the job or task at hand). His distribution of time-span-dependent organizational levels is likewise nonlinear. Indeed, he addresses heredity ("cognitive power"—see Figure 5), training, and task as elements of competence. This is analogous to Blanchard's transportable versus task skills.

Jaques's approach is quite cognitive in itself, seeming to display a Myers-Briggs propensity for the thinking function. Yet, later, in *Executive Leadership,* he does address the ramifications of the feeling function (MBTI "F") as well as the usefulness of intuition (MBTI "N") (Briggs-Myers and McCaulley, 1985). Thus, he is more inclusive than would first appear. He seems to be trying to compensate for others' overemphasis on the value of

charisma. His emphasis is more upon organizational structure and dynamics rather than on individual personality differences, which cannot be relied upon over either time or the breadth of the organization.

He even makes the seemingly enigmatic point (*Executive Leadership*, p. 180) that "clearly bounded general responsibilities paradoxically release initiative and creativity because the boundaries are clear. Unclear boundaries and lack of adequate limits always stifle initiative because people do not know how far they can push new ideas" (Jaques, 1991). While I suspect that this might be limited to ±2 or 3 σ (standard deviations), it is part of a strong argument for the value of organizational structure not only for the attainment of organizational goals, but also for the continued stability and health of its members. Jaques addresses the interdependence between workers and organizations as a long-term value.

"It is difficult to imagine a true leader who didn't need to manage or a true manager who didn't need to lead."

Jaques's longitudinal analyses of the depth of organizational talent (based upon the time-span potentials of company personnel) is most interesting. He has devised learning curves for time-span abilities such that he can project when specific individuals would be capable of assuming higher positions in the organization. While this might be difficult (if not impossible) to defend in court today, Jaques has certainly boldly gone where no one (to my knowledge) has gone before. His analytical, scientific approach lends credence to the term "management science," which otherwise seems a gross misnomer.

Nonetheless, he does not ignore the human factor, stating that "to be an effective managerial leader a person must really value the opportunity to work with subordinates and value being able to unleash their enthusiastic and effective collaboration" (1991, p. 72). To Jaques, there is no difference between a manager and a leader. It is difficult to imagine a true leader who didn't need to manage or a true manager who didn't need to lead. Such people (if they exist) are probably not in the mainstream of the industrial and governmental organizations with which we are mainly concerned.

NONLINEAR DEVELOPMENT

John D. Rockefeller said that "Good management consists of showing average people how to do the work of superior people" (Braude, 1961, p. 57). Perhaps such an approach includes the creation of an enabling organizational structure and work environment which place individuals in their appropriate level positions and addresses their future potentials. This is a major change from Peter Drucker's observation that "Most of what we call management consists of making it difficult for people to get their work done" (Albrecht & Zenke, 1985).

Even with the benefit of management training at the Defense Acquisition University and despite numerous acquisition reform initiatives, there has been little progress toward a new management paradigm. We need a meta management plan

rather than just program management training. It would, of necessity, include both leadership and supervision under its purview. According to Alfred North Whitehead, "The art of progress is to preserve order amid change, and to preserve change amid order" (Braude, 1961, p. 311). The qualitative nature of that order, I would suggest, differs for evolutionary, gradual change versus revolutionary Kuhnian, paradigmatic change. Transitioning from a binary or even linear model to a nonlinear, multidimensional one is such a change.

An example is in order. Employees today (for instance, in the federal government), obtain increases over present salaries via three avenues: merit raises and bonuses, cost of living increases, and promotions. The first is based on past history, generally of the preceding work year and against preset task objectives. The second does not involve work performance. But the third, promotions, are the case in point. A supervisory position is often filled by the "best qualified" worker—possibly a former subordinate to this position. In analyzing the candidates, what criteria are used? Work performance. But at what level? At the worker or subordinate level. Thus, an organization may lose its best worker by promoting him or her into a supervisory post for which he or she may not be suited. Using Blanchard's definition of competence, this is easily comprehended.

Jaques's approach is even more dramatically opposed to such a promotion schema. He analyzes the cognitive complexity of the job versus the cognitive power of the person and looks for a match. Such a schema depends on each candidate's future capabilities and performance rather than his or her past capabilities and performance.

Interestingly, Lawrence J. Peter popularized his "Peter Principle" based on people being promoted into positions beyond their competence to their "level of incompetence" (Peter, 1970, 1972). However, the similarity to Blanchard and Jaques is ephemeral. Peter's model was static. Once one reached one's level of incompetence, one stayed there forever. Not so with Jaques, who actually provides curves based upon natural human growth in ability over time. He shows that one naturally grows from level to level; he's much more optimistic than Peter. The growth is, however, gradual. It's also reminiscent of program management learning curves.

"[Jaques] analyzes the cognitive complexity of the job versus the cognitive power of the person and looks for a match."

Blanchard's model implies a similar understanding in that competence to him depends on innate abilities and his situational leadership techniques for developing task competence are recognized to have limits. No one can be competent at everything. Indeed, some early studies on programmed learning demonstrated that the hard-core unemployed could not necessarily be educated to perform given work tasks even with the very best instruction and equipment.

However, despite some nonlinearities such as Blanchard's normal (Gaussian)-shaped development curve and Jaques's discontinuous time-span levels of work performance, both still appear to predicate linear personal work development. In

other words, a worker proceeds through each level to attain higher levels. This parallels Peter's approach, which assumed that all the positions beyond the first level of incompetence were also levels of incompetence. I think this implies an assumption of evolutionary development that is not always applicable.

Activation or arousal theory (Figure 6) uses a nonlinear model for human performance, similar in shape (normal distribution) to Blanchard's situational leadership curve. According to activation theory, an individual's performance has an optimum point (local maximum). Starting from the left, as stimulation (e.g., task demands upon the employee) increases (along the *x*-axis), output increases (along the *y*-axis). There is, it may be noted, a point at which the curve becomes nearly linear. But the slope becomes zero at the "characteristic point" (the local maximum—highest point of the curve). After that, the curve slopes downwards, again resembling a straight line for a time, until it asymptotically approaches zero on the *y*-axis.

To the left of the "characteristic point," the employee is underutilized and, probably, bored. To the right of the "characteristic point," the employee is overutilized and, probably, overstressed. Output is nonoptimal under both these conditions. I would suggest that such a model could be applied not only to the quantity of work assigned, but also to the quality of work assigned. A person may have a lot of work to do, but if that work is well below the scope of the work of which that person is capable, that person will still be underutilized. Such a situation is analogous to Jaques's mismatch between a person's present time-span capability versus the time-span of his or her present work or position.

Another model may be illustrative of such a situation. The MBTI delineates 16 personal preference modes based upon Jungian psychology. Per Blanchard's

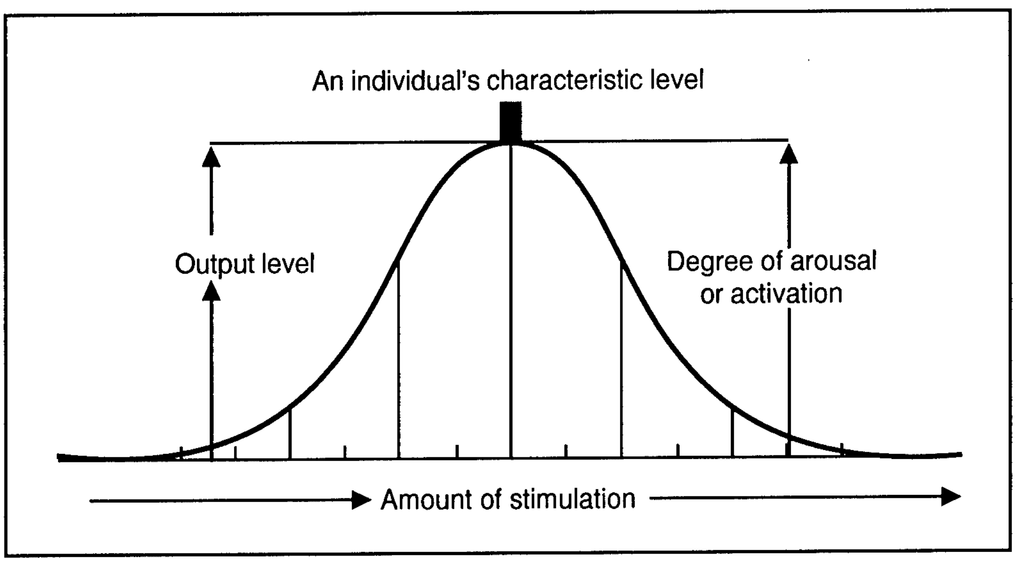


Figure 6. Activation or Arousal Theory

"leadership bridge" model, the Myers-Briggs can be correlated to work capabilities and performance as well as used in management and supervisory strategies. The Myers-Briggs is a nonlinear, nondevelopmental system, though it does indicate the possibility of consciously developing the ability for a person to operate in modes other than one's natural (preferred) one.

In analyzing executive abilities, the Industrial College of the Armed Forces (Knowlton & McGee, 1996) determined that certain of the preferences (especially the N or "iNtuitive" function as opposed to the S, or sensate, function) were more conducive to such abilities and gathered statistics on the distributions of such preferences and characteristics among corporate, government, and student populations. Others have shown, not surprisingly, that certain occupations have a far from statistically average distribution of MBTI preferences.

I participated in a small study in the Navy in which the engineers were predominantly from two specific preferences (of the 16). The vast majority were ISTJ or ESTJ (either introverted or extroverted, sensate, thinking, and judging). In Blanchard's leadership bridge (Good et al., 1992), SJs are called "preservers." As with all of the categories, they have positive and negative characteristic tendencies (which an individual can either accept or overcome through personal effort). Since SJs tend to build on the past and focus on the present (see Kiersey & Bates, 1984), they may find it difficult to adapt to the present Defense Department emphasis on personnel empowerment, out-of-the-box creativity, acquisition reform, open systems, and nonlinear thinking.

Thus, while I certainly do not dispute either of our authors' contentions that there is an appropriate level for each particular person on each particular task, I do question the implication that each individual must proceed linearly through and beyond each level of development or task from a job or position perspective (level of abstraction). This is akin to the situation in which a college may allow candidates to take placement examinations vice taking each and every prerequisite to higher learning. I do wonder, however, if perhaps, in identifying the levels in an entire corporation (with growth curves for present employees), Jaques has implied that he would hire personnel based upon their present and future growth curves as matched to the needs of the company—implying a nonlinear assignment based upon future performance and predicted growth. His extensive description of normal human development might merely be provided as the norm, without precluding deviations from that norm.

"Jaques has implied that he would hire personnel based upon their present and future growth curves as matched to the needs of the company...."

I would contend that there might very well be a correlation between Jaques's cognitive processes and developmental levels and the MBTI. His charts of developmental levels, for instance, take age into account such that two individuals might be at the same level despite great variances in chronological age. But then, the intelligence quotient purports to measure the ratio of cognitive age to chronological age. This reminds me of the old television

show "Doogie Howser, M.D." It is debatable, however, whether our society is ready for such differences between societal position and age.

Nonetheless, it is difficult to argue, at least in theory, with an attempt to match people to jobs based upon competency. But we seem to be making little progress in a practical vein. It may be, however, that by implementing structural and procedural improvements à la Blanchard and Jaques, many other enhancements might occur as well. Indeed, to evaluate the prospects of a proposed intervention, one must attempt to project the effects of the change upon the whole, not just upon a portion of the whole. There is simply too much interaction and interdependency within existing organizations to have any hope of performing a change and avoiding second-, third-, and higher- order ramifications. To understand what may happen, what is happening, and what will happen, we need a more comprehensive understanding of people and management. Thus, I applaud the innovative and valuable contributions of both Blanchard and Jaques toward a "unified field theory of management."

Neal Pollock is the executive assistant for engineering and program assessment to the Program Executive Officer for Space, Communications, and Sensors (PEO-SCS) and has served as the executive secretary to the Multifunctional Information Distribution System (MIDS) international Steering Committee since 1994.
(E-mail address: pollockn@spawar.navy.mil)

References

Albrecht, K., & Zenke, R. (1985). *Service America!* (p. 106). New York: Warner Books.

Blanchard, K., Zigarmi, P., & Zigarmi, D. (1985). *Leadership and the one minute manager.* New York: William Morrow.

Blanchard, K. (1991). *The one minute manager.* New York: William Morrow, NY.

Braude, J. (1961). *New treasury of stories for every speaking and writing occasion.* Englewood Cliffs, NJ: Prentice-Hall.

Briggs-Myers, I., & McCaulley, M. H. (1985). *Manual: A guide to the development and use of the Myers-Briggs type indicator.* Palo Alto, CA: Consulting Psychologists Press.

Cascio, W. (1978). *Applied psychology in personnel management.* Reston, VA: Reston Publishing (Prentice-Hall).

Frisch, F. (1995, Spring). Book reviews: The balanced budget. *Acquisition Review Quarterly,* 172–185.

Good, L., Hill, T., & Blanchard, K. (1992). *The leadership bridge participant workbook.* Escondido, CA: Blanchard Training and Development.

Jaques, E. (1991). *Executive leadership.* Arlington, VA: Cason Hall.

Jaques, E. (1989; 2nd ed. 1996). *Requisite organization.* Arlington, VA: Cason Hall.

Kiersey, D., & Bates, M. (1984). *Please understand me.* Del Mar, CA: Prometheus Nemesis.

Knowlton, W., & McGee, M. (1996). *Strategic leadership and personality: making the MBTI relevant* (Industrial College of the Armed Forces). Washington, DC: National Defense University Press.

Peter, L. J. (1972). *The Peter prescription.* New York: William Morrow.

Peter, L. J., & Hull, R. (1970, 1996). *The Peter principle.* New York: William Morrow.

Pritchett, P. (1993). *High velocity culture change.* Dallas, TX: Pritchett Publishing.

ACQUISITION REVIEW QUARTERLY GUIDELINES FOR CONTRIBUTORS

The *Acquisition Review Quarterly (ARQ)* is a scholarly peer-reviewed journal published by the Defense Acquisition University. All submissions receive a masked review to ensure impartial evaluation.

SUBMISSIONS

Submissions are welcomed from anyone involved in the Defense acquisition process. Defense acquisition is defined as the conceptualization, initiation, design, development, test, contracting, production, deployment, logistic support, modification, and disposal of weapons and other systems, supplies, or services to satisfy Defense Department needs, or intended for use in support of military missions.

RESEARCH ARTICLES

Manuscripts should reflect research or empirically-supported experience in one or more of the aforementioned areas of acquisition. Research or tutorial articles should not exceed 4,500 words. Opinion pieces should be limited to 1,500 words.

We publish Defense Acquisition research articles that involve systemic inquiry into a significant research question. The article must produce a new or revised theory of interest to the acquisition community. You must use a reliable, valid instrument to provide your measured outcomes.

MANUSCRIPT SECTIONS

The introduction should state the purpose of the article and concisely summarize the rationale for the undertaking.

The methods section should include a detailed methodology that clearly describes work performed. Although it is appropriate to refer to previous publications in this section, the author should provide enough information so that the experienced reader need not read earlier works to gain understanding of the methodology.

The results section should concisely summarize findings of the research and follow the train of thought established in the methods section. This section should not refer to previous publications, but should be devoted solely to the current findings of the author.

The discussion section should emphasize the major findings of the study and its significance. Information presented in the aforementioned sections should not be repeated.

RESEARCH CONSIDERATIONS

Contributors should also consider the following questions in reviewing their research-based articles prior to submission:

- Is the research question significant?
- Are research instruments reliable and valid?
- Are outcomes measured in a way clearly related to the variables under study?
- Does the research design fully and unambiguously test the hypothesis?
- Did you build needed controls into the study?

Contributors of research-based submissions are also reminded they should share any materials and methodology necessary to verify their conclusions.

CRITERIA FOR TUTORIALS

Tutorials should provide special instruction or knowledge relevant to an area of defense acquisition to inform the Defense Acquisition Workforce.

Topics for submissions should rely on or be derived from observation or experiment, rather than theory. The submission should provide knowledge in a particular area for a particular purpose.

OPINION CRITERIA

Opinion articles should reflect judgments based on the special knowledge of the expert. Opinion articles should be based on observable phenomena and presented in a factual manner; that is, submissions should imply detachment. The observation and judgment should not reflect the author's personal feelings or thoughts. Nevertheless, opinion pieces should clearly express a fresh point of view, rather than negatively criticize the view of another previous author.

MANUSCRIPT STYLE

We will require you to recast your last version of the manuscript, especially citations (e.g., footnotes or endnotes) into the format required in two specific style manuals. The *ARQ* follows the author (date) form of citation. We expect you to use the Publication Manual of the American Psychological Association (4th Edition), and the Chicago Manual of Style (14th Edition). The *ARQ* follows the author (date) form of citation.

Contributors are encouraged to seek the advice of a reference librarian in completing citations of government documents. Standard formulas of citations may give only incomplete information in reference to government works. Helpful guidance is also available in Garner, D.L. and Smith, D.H., 1993, The Complete Guide to Citing Government Documents: A Manual for Writers and Librarians (Rev. Ed.), Bethesda, MD: Congressional Information Service, Inc.

Copyright Information

The *ARQ* is a publication of the United States Government and as such is not copyrighted. Contributors of copyrighted works and copyright holders of works for hire are strongly encouraged to request that a copyright notification be placed on their published work as a safeguard against unintentional infringement. The work of federal employees undertaken as part of their official duties is not subject to copyright.

In citing the work of others, it is the contributor's responsibility to obtain permission from a copyright holder if the proposed use exceeds the fair use provisions of the law (see U.S. Government Printing Office, 1994, Circular 92: Copyright Law of the United States of America, p. 15, Washington, DC: Author). Contributors will be required to submit a copy of the written permission to the editor before publication.

Manuscript Format

Pages should be double-spaced and organized in the following order: title page, abstract, body, reference list, author's note (if any), and figures or tables. To ensure anonymity, each paper should be submitted with a separate page that includes the author(s)'s name(s) and complete address, and the paper should include the title, abstract, keywords, body, complete set of references, along with tables and figures at the end. Authors are reminded not to refer to themselves or to their own work directly in the paper. Figures or tables should not be inserted (or embedded, etc.) into the text, but segregated one to a page following the text. Articles must be printable within one issue and should not exceed 4,500 words for research or tutorials and 1,500 words for opinion pieces; articles will not be printed in parts or in a continuing series. If material is submitted on a computer diskette, each figure or table should be recorded in a separate, exportable file (i.e., a readable .eps file). For additional information on the preparation of figures or tables, see CBE Scientific Illustration Committee, 1988, Illustrating Science: Standards for Publication, Bethesda, MD: Council of Biology Editors, Inc. Please restructure briefing charts and slides to a look similar to those in previous issues of *ARQ*.

The author (or corresponding author in the case of multiple authorship) should attach to the manuscript a signed cover letter that provides the author's name, address, and telephone number (fax and Internet addresses are also appreciated). The letter should verify that the submission is an original product of the author; that it has not been published before; and that it is not under consideration by another publication. Details about the manuscript should also be included in this letter: for example, its title, word length, the need for copyright notification, the identification of copyrighted material for which permission must be obtained, a description of the computer application programs and file names used on enclosed diskettes, etc. A short biography of no more than 75 words and a photo of the author will be expected from each author. Author names and e-mail addresses are not part of the 75-word count.

The letter, one copy of the printed manuscript, and any diskettes should be

sturdily packaged and mailed to: Defense Systems Management College, Attn: DSMC Press (*ARQ*), 9820 Belvoir Road, Suite 3, Fort Belvoir, VA 22060-5565.

In most cases, the author will be notified that the submission has been received within 48 hours of its arrival. Following an initial review, submissions will be referred to referees and subsequent consideration by the *ARQ* Editorial Board.

Contributors may direct their questions to the Editor, *ARQ*, at the address shown above, by calling (703) 805-4290 (fax 805- 2917), or via the Internet at:

gonzalezd@dsmc.dsm.mil

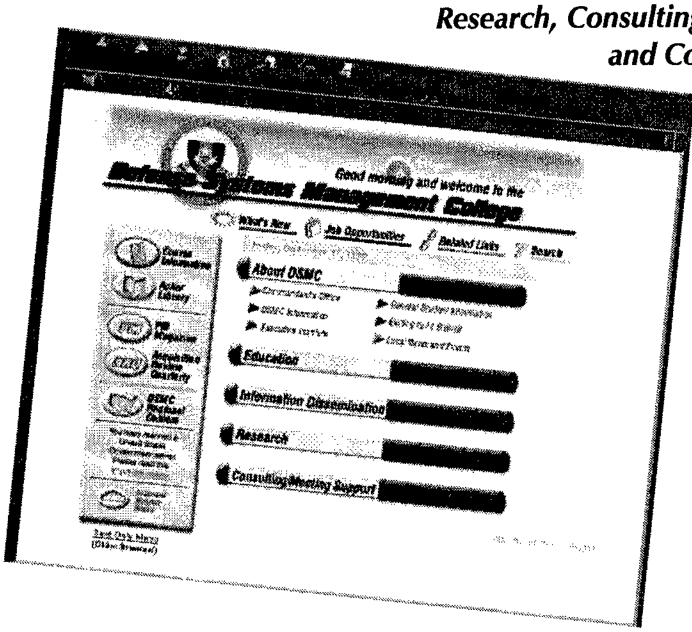
The DSMC Home Page can be accessed at:

http://www.dsmc.dsm.mil

DSMC'S Home Page

http://www.dsmc.dsm.mil

Your Online Access to Acquisition Research, Consulting, Information, and Course Offerings



Front Page
- Course Information
- David D. Acker Library
- PM Magazine
- ARQ Magazine
- What's New
- Acquistion Related Links

About DSMC
- Commandant's office
- DSMC Information
- Executive Institute
- General Student Information
- Getting to Ft. Belvoir
- Local News and Events

Education
- Acquisition Reform Learning Module
- Continuing Education
- Distance Learning
- DCAS
- DSMC Alumni Association (DSMCAA)
- DSMC Course List
- DSMC Divisions
- Education Opportunities
- EPMC
- EPMC Extranet
- Faculty Departments
- General Student Information
- IEVMC
- Individual Learning Program
- International Acquistion
- Learning Resource Center
- Past Performance
- Program Management
- Regional Centers
- Registrar

Information Dissemination
- Acquisition Events
- Best Practices
- Lessons Learned
- Links to Related Sites
- Manufacturing Resources
- Past Performance
- Publications
- Services

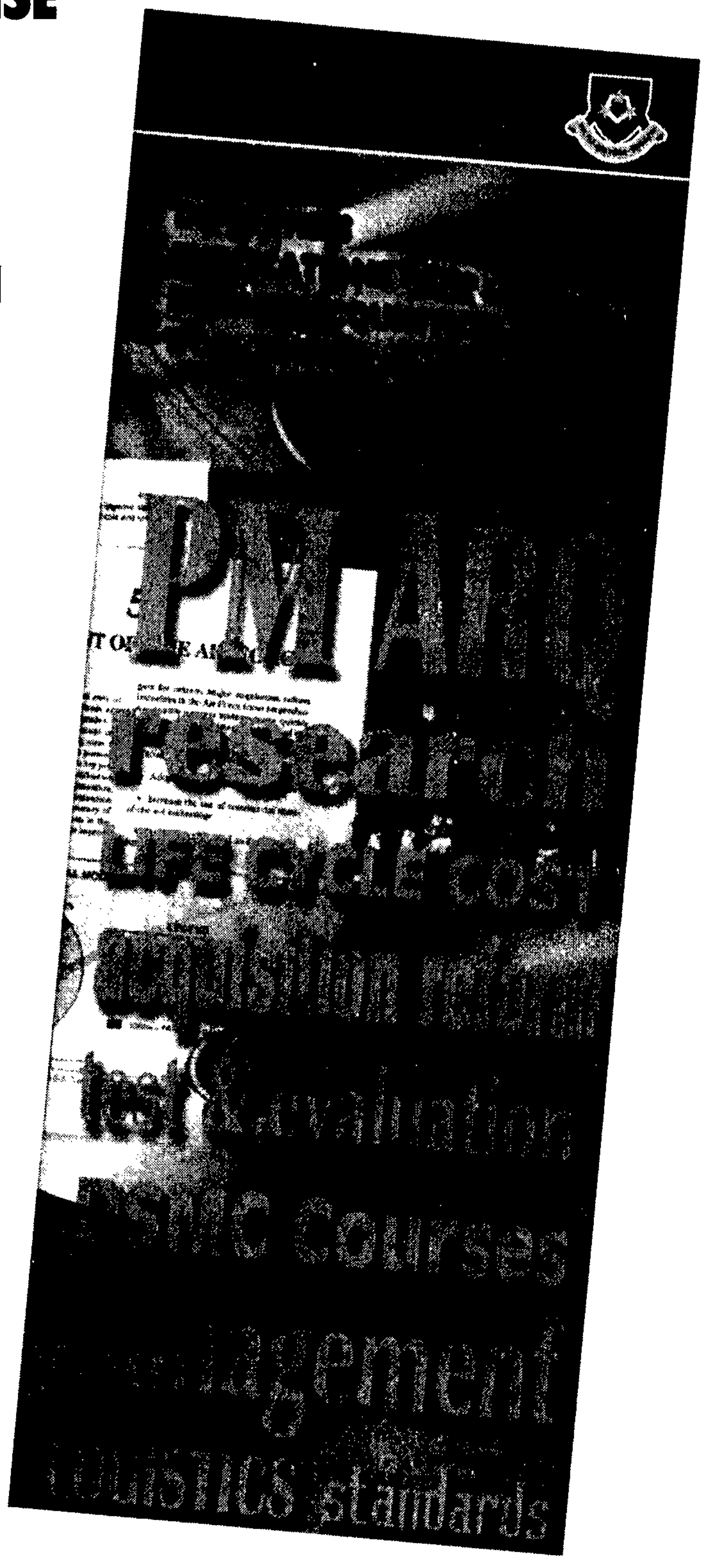
Research
- Acquisition Events
- Best Practices
- Lessons Learned
- DSMC Military Research Fellows
- Overview/Projects
- ROAR

Consulting/Meeting Support
- Consulting Services
- MDC/Group Services
- Acquisition Reform Learning Model

Now you can search the DSMC Website and our on-line publications!

DSMC PRESS PUBLICATIONS FOR THE DEFENSE ACQUISITION WORKFORCE

Newly revised and updated, this free brochure is yours by faxing a request to the DSMC Press: (703) 805-2917 or DSN 655-2917. The brochure lists the publications offered by and through the College, including titles, abstracts, prices, sources, and reference numbers.



PM/ARQ SUBSCRIPTIONS

FREE SUBSCRIPTIONS

☐ ***PROGRAM MANAGER*** **MAGAZINE (*PM*)**

NOW FREE TO EVERYONE

☐ ***ACQUISITION REVIEW QUARTERLY* (*ARQ*)**

Name and Title (Please Print)

Organization

Address

( )

City | State | Zip | Daytime Phone

Work E-Mail | Home E-Mail

Date Signed

CHANGE OF ADDRESS

☐ ***PROGRAM MANAGER*** **MAGAZINE (*PM*)**

☐ ***ACQUISITION REVIEW QUARTERLY* (*ARQ*)**

Name and Title (Please Print)

Organization

Address

( )

City | State | Zip | Daytime Phone

Work E-Mail | Home E-Mail

Date Signed